TITLE

METHOD TO PRODUCE PARA-HYDROXYBENZOIC ACID IN THE STEM TISSUE OF GREEN PLANTS BY USING A TISSUE-SPECIFIC PROMOTER

FIELD OF THE INVENTION

The invention relates to the fields of plant gene expression, molecular biology, and microbiology.

BACKGROUND OF THE INVENTION

Recent advances in genetic engineering have enabled the development of new biological platforms for the production of molecules, heretofore only synthesized by chemical routes. Although microbial fermentation is routinely exploited to produce of small molecules and proteins of industrial and/or pharmaceutical importance (antibiotics, enzymes, vaccines, etc.), the possibility of using green plants to manufacture a high volume of chemicals and materials has become an increasingly attractive alternative.

Using green plants to produce large amounts of compounds has two significant advantages over traditional chemical synthesis. First, green plants constitute a renewable energy source, as opposed to finite petrochemical resources. Because of photosynthesis, the only raw materials that are required to produce carbon-based compounds in green plants are carbon dioxide, water, and soil. Sunlight is the ultimate source of energy. Second, in comparison to existing fermentation facilities that are expensive and limited in size, green plants constitute a huge available biomass that could easily accommodate the large amounts of chemicals that are required for certain high-volume, low-cost applications.

Producing para-hydroxybenzoic acid in green plants transformed with 4-hydroxycinnamoyl-CoA hydratase/lyase (HCHL) has been previously described (Mayer *et al.*, *Plant Cell*, 13:1669-1682 (2001) and US SN 10/359369). Mitra *et al.* (*PLANTA*, 215:79-89 (2002)) express an HCHL in hairy root cultures of *Datura stramonium*. Expression of HCHL enzymes in plant cells leads to production of para-hydroxybenzoic acid (pHBA) from 4-coumaroyl-CoA (pHCACoA). The pHBA produced in plants is rapidly glucosylated by one or more endogenous UDP-glucosyltransferases into pHBA glucosides (both phenolic and ester glucosides) (Mayer *et al.*, *supra*; Mitra *et al.*, *supra*, and US SN 10/359369) that are subsequently sequestered in the plants' vacuoles.

pHCACoA is normally used by plants to make molecules that are secondary metabolites with roles as plant growth regulators, UV protectants, or cell wall components such as lignin, cutin, or suberin. Examples of secondary metabolites made from pHCACoA include caffeoyl-CoA and feruloyl-CoA.

Expression of *HCHL* genes in tobacco plants under the control of a constitutive promoter (*CaMV35S*) leads to plant growth defects such as interveinal leaf chlorosis, stunting, low pollen production, and male sterility (Mayer *et al.*, *supra*). As a result of constitutive HCHL expression (in all plant tissues), pHCACoA levels were depleted to a point where molecules derived from pHCACoA that are essential for plant growth and reproduction were no longer produced in adequate amounts.

HCHL expression needs to be targeted to cells where suitable pools of pHCACoA exist and where conversion to pHBA does not detrimentally affect plant growth and reproduction. Plant stem tissue contains a significant pool of available pHCACoA and can accommodate large fluxes to the phenylpropanoid pathway. In order to exploit the available substrate pool without causing detrimental effects to the plant, HCHL expression needs to be limited to plant stem tissue. In addition, expression levels need to be high enough to produce suitable quantities of pHBA. Robust tissue-specific plant promoters, namely those which are known to drive genes involved in cell wall biosynthesis, represent an attractive group of candidate promoters for HCHL expression.

Genes involved in the production of phenylpropanoid derivatives used in plant cell wall biosynthesis (which are expected to show a tissue-specific expression pattern) represent a source of possible promoters to drive tissue-specific HCHL expression. Examples of these genes include cinnamate-4-hydroxylase (*C4H*; GenBank® U71080), 4-coumaroyl-Coenzyme A ligase (*4CL1*; GenBank® U18675), para-coumarate 3-hydroxylase (*C3'H*; AC011765), and the genes encoding proteins responsible for the catalytic activity of cellulose synthase (*IRX1*, *IRX3*, *IRX5*, and their respective orthologs from rice and maize)(Taylor *et al.*, *PNAS*, 100(3):1450-1455 (2003)). Given the requirement that HCHL expression must be limited to stem tissue, it is unknown if any of these promoters are suitable for stem-specific expression. Use of these promoters for HCHL expression in plant stalk tissue has not been reported.

Cellulose is a polymer of β(1,4)-linked glucose. It is an essential component of both the primary and secondary cell walls in higher plants. Cellulose can make up to 90% of the dry weight of the secondary walls. In the plant cell wall, individual cellulose chains crystallize to form microfibrils. Cells involved in synthesizing the cellulose for the secondary cell wall represent an attractive target for tissue-specific expression of HCHL.

Cellulose synthesis is believed to involve a multienzyme complex situated at the plasma membrane (Taylor *et al.*, *Plant Cell*, 11(5):769-779 (1999); Taylor *et al.*, *supra* (2003)). Many of the cellulose synthase genes "*CesA* genes" are

classified as such based on highly-conserved motifs (Richmond and Sommerville, *Plant Physiol.*, 124:495-498 (2000) and Delmer, DP, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 50:245-276 (1999)). Many of the genes share homology with one another, yet appear to have different roles in cellulose biosynthesis. The CesA genes are a subset of a larger family of related genes which share some homology to one another. These genes form a family of cellulose synthase-like genes ("*csl*" genes; Taylor *et al.*, *supra* (2003); Richmond, T., *Genome Biol.*, 1(4):reviews 3001.1-3001.6 (2000)) whose exact function is not known.

Use of promoters from CesA genes have previously been described. Turner *et al.* (WO 00/070058) describe the use of cellulose synthase genes or promoters (*IRX3*) for modulating enzymes involved in the synthesis of plant cell walls. Jones *et al.* (*Plant Journal*, 26(2):205-216 (2001)) described the utility of the *IRX3* promoter to down-regulate genes involved with lignin synthesis in plant stalk tissue. Allen *et al.* (WO 00/04166) describe methods related to altering cellulose synthase genes (*CesA*). Stalker *et al.* (WO 98/18949) describe a *CesA* homolog from cotton (*Gossypium hirsutem*) and methods associated with altering cotton fiber and wood quality. Arioli *et al.* (WO 98/00549) describe methods for manipulating a cellulose synthase-like gene (*rsw1*) for altering cellulose biosynthetic properties. None of these references teach the use of a cellulose synthase-like gene promoter to drive *HCHL* expression.

The *IRX3* gene was putatively identified as encoding the cellulose synthase catalytic subunit from Arabidopsis (Turner *et al.*, *Plant Cell*, 9(5): 689-701 (1997). Expression of the *IRX3* gene was shown to be normally limited to plant stem tissue as no detectable mRNA transcript was measured in leaf tissue (Taylor *et al.*, *supra* (1999)). It was later reported that the catalytic activity of cellulose biosynthesis is attributed to a multi-subunit complex formed by the proteins encoded by the *IRX1*, *IRX3*, and *IRX5* genes (Taylor *et al.*, *Plant Cell*, 12:2529-2539 (2000) and Taylor *et al.*, *supra* (2003)). These three genes identified from Arabidopsis show essentially the same expression patterns. Expression of these genes is normally limited to cells involved in secondary cell wall biosynthesis. Additionally, orthologs of these genes may exhibit similar tissue-specific expression patterns, namely expression in cells that produce cellulose for secondary cell wall synthesis. The prior art does not teach use of the promoters from *IRX1*, *IRX3*, or *IRX5* (or orthologs thereof) for stem tissue expression of HCHL.

The problem to be solved is to identify regulatory sequences that allow targeted HCHL expression in plant tissues where significant pHBA accumulation

can occur without adversely affecting the synthesis of compounds essential for plant growth and development. In other words, technology needs to be developed that allows for HCHL-mediated pHBA production in plants without negative effects on plant performance in the field.

SUMMARY OF THE INVENTION

Methods and materials are presented for the production of para-hydroxybenzoic acid in genetically modified green plants by selectively expressing hydroxycinnamoyl CoA hydratase/lyase genes using tissue-specific promoters. The promoters from the genes involved in the formation of the cellulose synthase catalytic complex are suitable for tissue-specific expression of HCHL in plants. The promoters from *Arabidopsis thaliana* genes *AtCesA4* (*IRX5*), *AtCesA7* (*IRX3*), and *AtCesA8* (*IRX1*) are suitable for tissue-specific expression of HCHL. Additionally, the promoters of orthologous genes from maize and rice are also suitable for stem tissue targeted expression of HCHL.

The invention embodies a method to selectively produce para-hydroxybenzoic acid in plant stem tissue comprising:

a) growing a plant under suitable conditions, the plant comprising
   i) an endogenous source of para-coumaroyl-CoA;
   ii) a 4-hydroxycinnamoyl-CoA hydratase/lyase (HCHL) expression cassette comprising a tissue-specific promoter isolated from a cellulose synthase gene encoding a protein involved in the formation of a cellulose synthesis catalytic complex, wherein said cellulose synthesis catalytic complex catalyzes cellulose synthesis in secondary cell wall formation in plant vascular tissue, said tissue-specific promoter operably linked to a nucleic acid molecule encoding a 4-hydroxycinnamoyl-CoA hydratase/lyase enzyme; and
   iii) a gene encoding a para-hydroxybenzoic acid UDP-glucosyltransferase;
b) recovering unconjugated para-hydroxybenzoic acid and para-hydroxybenzoic acid glucoside from the plant;
c) hydrolyzing para-hydroxybenzoic acid glucoside; and
d) recovering unconjugated para-hydroxybenzoic acid.

The tissue-specific promoter is selected from the group consisting of SEQ ID Nos:26, 43, 44, 45, 46, 49, 81, 82, and 83. The HCHL expression cassette is represented by SEQ ID NO:30. The nucleic acid molecule encoding HCHL is isolated from a bacterium selected from the group consisting of Pseudomonas, Caulobacter, Delftia, Sphingomonas, and Amycolatopsis. The bacteria from which the nucleic acid is isolated is selected from the group consisting of

Pseudomonas putida (DSM 12585), *Pseudomonas fluorescens* AN103, *Pseudomonas putida* WCS358, *Pseudomonas sp.* HR199, *Delftia acidovorans*, *Amycolatopsis sp.* HR167, *Sphingomonas paucimobilis*, and *Caulobacter crescentus*.

The nucleic acid molecule encoding HCHL is selected from the group consisting of SEQ ID NO:5, 58, 59, 60, 62, 63, and 64. The nucleic acid molecule encoding HCHL encodes the polypeptide of SEQ ID 61. The nucleic acid molecule encoding HCHL coding is isolated from *Psuedomonas putida* DSM 12585. The nucleic acid molecule encoding HCHL encodes the polypeptide of SEQ ID NO:6. The nucleic acid molecule encoding HCHL is SEQ ID NO:5. The gene encoding the para-hydroxybenzoic acid UDP-glucosyltransferase may be endogenous or exogenous to the plant. The gene encoding para-hydroxybenzoic acid UDP-glucosyltransferase is selected from the group consisting of SEQ ID NOs:65, 66, and 67 and is recombinantly expressed in the plant whereby para-hydroxybenzoic acid glucose ester is selectively produced. The tissue-specific promoter of said HCHL expression cassette preferentially expresses active HCHL in said plant stem tissue at levels at least ten, times higher than expression levels measured in leaf tissue of said plant. More preferred embodiments show preferential expression levels of active HCHL in said plant stem tissue of 20 times to 50 times greater than expression levels measured in the leaf tissue of the plant.

Another method to selectively produce para-hydroxybenzoic acid in plant stem tissue comprises

a) Providing a plant comprising
    i. an endogenous source of para-coumaroyl-CoA;
    ii. a 4-hydroxycinnamoyl-CoA hydratase/lyase (HCHL) expression cassette comprising a tissue-specific promoter isolated from a cellulose synthase gene encoding a protein involved in the formation of the cellulose synthesis catalytic complex, the tissue-specific promoter operably linked to a nucleic acid molecule encoding a 4-hydroxycinnamoyl-CoA hydratase/lyase enzyme from *Caulobacter crescentus* having at least 50% higher catalytic efficiency in converting para-hydroxycinnamoyl-CoA to para-hydroxybenzoic acid in comparison to catalystic efficienty of an HCHL enzyme from *Psuedomonas putida* or *Pseudomonas fluorescens* expressed under similar conditions; wherein said cellulose

synthesis catalytic complex catalyzes cellulose synthesis in secondary cell wall formation in plant vascular tissue; and

iii. a gene encoding a para-hydroxybenzoic acid UDP-glucosyltransferase;

b. growing a plant under suitable conditions whereby unconjugated para-hydroxybenzoic acid and para-hydroxybenzoic acid glucosides are produced;

c. recovering unconjugated para-hydroxybenzoic acid and para-hydroxybenzoic acid glucoside from the plant;

d. hydrolyzing para-hydroxybenzoic acid glucoside; and

e. recovering unconjugated para-hydroxybenzoic acid.

The nucleic acid molecule used in this method encodes an amino acid sequence as provided by SEQ ID NO:61. The plant is selected from the group consisting of tobacco, *Arabidopsis,* sugar beet, sugar cane, soybean, rapeseed, sunflower, cotton, corn, alfalfa, wheat, barley, oats, sorghum, rice, canola, millet, beans, peas, rye, flax, and forage grasses. The tissue-specific promoter is isolated from a gene selected from the group consisting of: *AtCesA4 (IRX5), AtCesA7 (IRX3), AtCesA8 (IRX1), ZmCesA10, ZmCesA11, ZmCesA12,* the *Oryza savita* (japonica cultivar) ortholog of *ZmCesA10,* the *Oryza savita* (japonica cultivar) ortholog of *ZmCesA11,* and the *Oryza savita* (japonica cultivar) ortholog of *ZmCesA12.*

The tissue-specific promoter is selected from the group consisting of SEQ ID NOs:26, 43, 44, 45, 46, 49, 81, 82, and 83. The gene encoding para-hydroxybenzoic acid UDP-glucosyltransferase may be endogenous or exogenous to the plant and is recombinantly expressed in the plant whereby para-hydroxybenzoic acid glucose ester is selectively produced. The gene encoding para-hydroxybenzoic acid UDP-glucosyltransferase is selected from the group consisting of SEQ ID NOs:65, 66, and 67.

BRIEF DESCRIPTION OF THE FIGURES AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the sequence listing, the Figures, and the detailed description that together form this application.

Figure 1 shows the enzyme pathway to produce pHBA in transgenic plants. The HCHL enzyme converts 4-coumaroyl-CoA to pHBA in the cytosol. A pHBA UDP-glucosyltransferase glucosylates the pHBA to produce a pHBA glucoside. The pHBA glucoside is subsequently stored and accumulated in the plant's vacuoles.

Figure 2 shows Michaelis-Menten and Wolf-Augustinsson-Hofstee plots illustrating kinetic properties of the recombinantly produced, purified HCHL enzyme of *Pseudomonas putida* (DSM 12585).

Figure 3 shows the linear relationship between HCHL activity and pHBA production in stalk tissue of transgenic lines expressing the HCHL gene of *Pseudomonas putida* (DSM 12585).

Figure 4 shows an unrooted single most parsimonious tree of the CesA proteins from maize and *Arabidopsis* found by the Branch and Bound algorithm of the PAUP program. (Swofford, DL, PAUP*: Phylogenetic analysis using parsimony (and other methods), Volume Version 4 (Sinauer Associates, Sunderland, MA)). Branch lengths are proportionate to the inferred number of amino acid substitutions, which are shown in bold font. Bootstrap values (%) supporting the monophyletic groups are shown along the branches in parentheses. *Arabidopsis* CesA protein sequences were deduced from the publicly available GenBank® nucleotide sequence (Table 7). (See also Example 4.)

Figure 5: Expression of the maize *CesA* genes in different tissues as compiled from the Massively Parallel Signature Sequencing (MPSS) database (Brenner *et al.*, *Proc. Natl. Acad. Sci. USA*, 97(4):1665-1670 (2000); Brenner *et al.*, *Nat. Biotech.*, 18:630-634 (2000); Hoth *et al., J. Cell. Sci.,* 115:4891-4900 (2002); Meyers *et al., Plant J.*, 32:77-92 (2002); US 6,265,163; and US 6,511,802). A comparison of stem versus leaf tissue expression was tabulated from the expression data (See also Example 5, and Table 9).

Figure 6 shows a phylogenetic tree produced by CLUSTAL W of putative and *bona fide* HCHL enzymes identified from a BLAST search of public databases.

Figure 7 shows the Michaelis-Menten plot illustrating the kinetic properties of recombinantly produced HCHL enzymes from *Caulobacter crescentus*, *Pseudomonas putida* (DSM12585), and *Pseudomonas fluorescens* AN103.

The following 83 sequence descriptions and sequences listings attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825. The Sequence Descriptions contain the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IYUB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for

nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NO:1 is the nucleic acid sequence of the 5' primer (Primer 1) useful for amplifying the *4CL-1* open reading frame (ORF) from *Arabidopsis thaliana* and its cloning into the *E. coli* expression vector pET28a.

SEQ ID NO:2 is the nucleic acid sequence of the 3' primer (Primer 2) useful for amplifying the *4CL-1* ORF of *Arabidopsis thaliana* and its cloning into the *E. coli* expression vector pET28a.

SEQ ID NO:3 is the nucleic acid sequence of the 5' primer (Primer 3) useful for amplifying the *HCHL* gene of *Pseudomonas putida* (DSM 12585) from genomic DNA of this organism.

SEQ ID NO:4 is the nucleic acid sequence of the 3' primer (Primer 4) useful for amplifying the *HCHL* gene of *Pseudomonas putida* (DSM 12585) from genomic DNA of this organism.

SEQ ID NO:5 is the nucleic acid sequence of the *HCHL* coding sequence from *Pseudomonas putida* (DSM 12585)

SEQ ID NO:6 is the deduced amino acid sequence of the HCHL protein of *Pseudomonas putida* (DSM 12585)

SEQ ID NO:7 is the nucleic acid sequence of the 5' primer (Primer 5) useful for amplifying the *HCHL* coding sequence from *Pseudomonas putida* (DSM 12585) and its cloning into the *E. coli* expression vector pET29a.

SEQ ID NO:8 is the nucleic acid sequence of the 3' primer (Primer 6) useful for amplifying the *HCHL* coding sequence from *Pseudomonas putida* (DSM 12585) and its cloning into the *E. coli* expression vector pET29a.

SEQ ID NO:9 is the nucleic acid sequence of another 3' primer (Primer 7) useful for amplifying the *HCHL* ORF from *Pseudomonas putida* (DSM 12585) flanked by *Nde*I and *Hind*III restriction sites and its cloning into the *E. coli* expression vector pET29a.

SEQ ID NO:10 is the amino acid sequence of a variant of the HCHL protein expressed from pET29a carrying a hexa-histidine tag.

SEQ ID NO:11 is the nucleic acid sequence of the 5' primer (Primer 8) useful for amplifying the promoter from the *ACTIN2* gene of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:12 is the nucleic acid sequence of the 3' primer (Primer 9) useful for amplifying the promoter from the *ACTIN2* gene of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:13 is the nucleic acid sequence of the *ACTIN2* promoter used by applicants for expression of the *HCHL* coding sequence in plants.

SEQ ID NO:14 is the nucleic acid sequence of another 5' primer (Primer 10) useful for amplifying the *HCHL* coding sequence of *Pseudomonas putida* (DSM 12585) that introduces a *Pag*I restriction site at the start codon of the gene.

SEQ ID NO:15 is the nucleic acid sequence of the 5' primer (Primer 11) useful for amplifying the *C4H* promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:16 is the nucleic acid sequence of the 3' primer (Primer 12) useful for amplifying the *C4H* promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:17 is the nucleic acid sequence of the *C4H* promoter of *Arabidopsis thaliana*.

SEQ ID NO:18 is the nucleic acid sequence of the 5' primer (Primer 13) useful for amplifying the *4CL-1* promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:19 is the nucleic acid sequence of the 3' primer (Primer 14) useful for amplifying the *4CL-1* promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:20 is the nucleic acid sequence of the *4CL-1* promoter of *Arabidopsis thaliana*.

SEQ ID NO:21 is the nucleic acid sequence of the 5' primer (Primer 15) useful for amplifying the *C3'H* promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:22 is the nucleic acid sequence of the 3' primer (Primer 16) useful for amplifying the *C3'H* promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:23 is the nucleic acid sequence of the *C3'H* promoter of *Arabidopsis thaliana*.

SEQ ID NO:24 is the nucleic acid sequence of the 5' primer (Primer 17) useful for amplifying the *AtCesA7* (*IRX3*) promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:25 is the nucleic acid sequence of the 3' primer (Primer 18) useful for amplifying the *AtCesA7* (*IRX3*) promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:26 is the nucleic acid sequence of the *AtCesA7 (IRX3)* stem-specific promoter of *Arabidopsis thaliana*.

SEQ ID NO:27 is the nucleic acid sequence of the *C4H* promoter fused to the *HCHL* coding sequence of *Pseudomonas putida* (DSM 12585).

SEQ ID NO:28 is the nucleic acid sequence of the *4CL-1* promoter fused to the *HCHL* coding sequence of *Pseudomonas putida* (DSM 12585).

SEQ ID NO:29 is the nucleic acid sequence of the *C3'H* promoter fused to the *HCHL* coding sequence of *Pseudomonas putida* (DSM 12585).

SEQ ID NO:30 is the nucleic acid sequence of the *AtCesA7* (*IRX3*) promoter fused to the *HCHL* coding sequence of *Pseudomonas putida* (DSM 12585).

SEQ ID NO:31 is the nucleic acid sequence of the *ZmCesA10* gene coding sequence (GenBank® Accession No. AY372244).

SEQ ID NO:32 is the deduced amino acid sequence of the ZmCesA10 enzyme.

SEQ ID NO:33 is the nucleic acid sequence of the *ZmCesA11* gene coding sequence (GenBank® Accession No. AF372245).

SEQ ID NO:34 is the deduced amino acid sequence of the ZmCesA11 enzyme.

SEQ ID NO:35 is the nucleic acid sequence of the *ZmCesA12* gene coding sequence (GenBank® Accession No. AF372246).

SEQ ID NO:36 is the deduced amino acid sequence of the ZmCesA12 enzyme.

SEQ ID NO:37 is the nucleic acid sequence of the rice gene identified as the ortholog to the *ZmCesA10* gene.

SEQ ID NO:38 is the deduced amino acid sequence of the rice gene identified as the ortholog to the *ZmCesA10* gene.

SEQ ID NO:39 is the nucleic acid sequence of the rice gene identified as the ortholog to the *ZmCesA11* gene.

SEQ ID NO:40 is the deduced amino acid sequence of the rice gene identified as the ortholog to the *ZmCesA11* gene.

SEQ ID NO:41 is the nucleic acid sequence of the rice gene identified as the ortholog to the *ZmCesA12* gene.

SEQ ID NO:42 is the deduced amino acid sequence of the rice gene identified as the ortholog to the *ZmCesA12* gene.

SEQ ID NO:43 is the nucleic acid sequence of the 2500 nucleotide bp 5' to the start codon of the rice gene orthologous to *ZmCesA10* considered to be a rice promoter useful for driving stem tissue-specific HCHL expression.

SEQ ID NO:44 is the nucleic acid sequence of the 2500 nucleotide bp 5' to the start codon of the rice gene orthologous to *ZmCesA11* considered to be a rice promoter useful for driving stem tissue-specific HCHL expression.

SEQ ID NO:45 is the nucleic acid sequence of the 2500 nucleotide bp 5' to the start codon of the rice gene orthologous to *ZmCesA12* considered to be a rice promoter useful for driving stem tissue-specific HCHL expression.

SEQ ID NO:46 is the nucleic acid sequence of the *Arabidopsis AtCesA4* (*IRX5*) stem-specific promoter.

SEQ ID NO:47 is the nucleic acid sequence of the 5' primer (Primer 19) useful for amplifying the *AtCesA4* (*IRX5*) promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:48 is the nucleic acid sequence of the 3' primer (Primer 20) useful for amplifying the *AtCesA4* (*IRX5*) promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:49 is the nucleic acid sequence of the *Arabidopsis AtCesA8* (*IRX1*) stem-specific promoter.

SEQ ID NO:50 is the nucleic acid sequence of the 5' primer (Primer 21) useful for amplifying the *AtCesA8* (*IRX1*) promoter of *Arabidopsis thaliana* from genomic DNA of this organism.

SEQ ID NO:51 is the nucleic acid sequence of the 3' primer (Primer 22) useful for amplifying the *AtCesA8* (*IRX1*) promoter of *Arabidopsis thaliana* from genomic DNA of this organism

SEQ ID NO:52 is the nucleic acid sequence of the first member of a primer pair (Primer 23) used to amplify the promoter of the rice gene identified as the ortholog of the *ZmCesA10* gene.

SEQ ID NO:53 is the nucleic acid sequence of the second member of a primer pair (Primer 24) used to amplify the promoter of the rice gene identified as the ortholog of the *ZmCesA10* gene.

SEQ ID NO:54 is the nucleic acid sequence of the first member of a primer pair (Primer 25) used to amplify the promoter of the rice gene identified as the ortholog of the *ZmCesA11* gene.

SEQ ID NO:55 is the nucleic acid sequence of the second member of a primter pair (Primer 26) used to amplify the promoter of the rice gene identified as the ortholog of the *ZmCesA11* gene.

SEQ ID NO:56 is the nucleic acid sequence of the first member of a primer pair (Primer 27) used to amplify the promoter of the rice gene identified as the ortholog of the *ZmCesA12* gene.

SEQ ID NO:57 is the nucleic acid sequence of the second member of a primer pair (Primer 28) used to amplify the promoter of the rice gene identified as the ortholog of the *ZmCesA12* gene.

SEQ ID NO:58 is the nucleic acid sequence of an *HCHL* gene from *Psuedomonas fluorescens* AN103 (GenBank® Accession No. Y13067).

SEQ ID NO:59 is the nucleic acid sequence of an *HCHL* gene from *Pseudomonas putida* WCS358 (GenBank® Accession No. Y14772).

SEQ ID NO:60 is the nucleic acid sequence of the coding sequence of an *HCHL* gene from *Caulobacter crescentus*.

SEQ ID NO:61 is the deduced amino acid sequence of the HCHL polypeptide from *Caulobacter crescentus*.

SEQ ID NO:62 is the nucleic acid sequence of an HCHL gene from *Pseudomonas* sp. HR199 (GenBank® Accession No. Y11520.1).

SEQ ID NO:63 is the nucleic acid sequence of an HCHL gene from *Delftia acidovorans* (GenBank® Accession No. AJ300832).

SEQ ID NO:64 is the nucleic acid sequence of an HCHL gene from *Amycolatopsis* sp. HR167 (GenBank® Accession No. AJ290449).

SEQ ID NO:65 is the nucleic acid sequence of a pHBA UDP-glucosyltransferase isolated from grape (*Vitis* sp.; US SN 10/359369).

SEQ ID NO:66 is the nucleic acid sequence of a pHBA UDP-glucosyltransferase isolated from *Eucalyptus grandis* (US SN 10/359369).

SEQ ID NO:67 is the nucleic acid sequence of a pHBA UDP-glucosyltransferase isolated from *Citrus mitis* (US SN 10/359369).

SEQ ID NO:68 is the nucleic acid sequence of a primer (Primer 29) used to amplify an *HCHL* ORF from *Caulobacter crescentus*.

SEQ ID NO:69 is the nucleic acid sequence of a primer (Primer 30) used to amplify the *HCHL* ORF from *Caulobacter crescentus*.

SEQ ID NO:70 is the nucleic acid sequence of a primer (Primer 31) used to amplify the *HCHL* ORF from *Pseudomonas fluorescens* AN103.

SEQ ID NO:71 is the nucleic acid sequence of a primer (Primer 32) used to amplify the *HCHL* ORF from *Pseudomonas fluorescens* AN103.

SEQ ID NO:72 is the nucleic acid sequence of a primer (Primer 33) used to amplify the *ACTIN2* gene from *Arabidopsis thaliana* for real time PCR analysis.

SEQ ID NO:73 is the nucleic acid sequence of a primer (Primer 34) used to amplify the *ACTIN2* gene from *Arabidopsis thaliana* for real time PCR analysis.

SEQ ID NO:74 is the nucleic acid sequence of a primer (Primer 35) used as a probe for the *ACTIN2* gene from *Arabidopsis thaliana* for real time PCR analysis.

SEQ ID NO:75 is the nucleic acid sequence of a primer (Primer 36) used to amplify the Caulobacter *HCHL* gene during real time PCR analysis.

SEQ ID NO:76 is the nucleic acid sequence of a primer (Primer 37) used to amplify the Caulobacter *HCHL* gene during real time PCR analysis.

SEQ ID NO:77 is the nucleic acid sequence of a primer (Primer 38) used as a probe for the Caulobacter *HCHL* gene during real time PCR analysis.

SEQ ID NO:78 is the nucleic acid sequence of a primer (Primer 39) used to amplify the Pseudomonas *HCHL* gene during real time PCR analysis.

SEQ ID NO:79 is the nucleic acid sequence of a primer (Primer 40) used to amplify the Pseudomonas *HCHL* gene during real time PCR analysis.

SEQ ID NO:80 is the nucleic acid sequence of a primer (Primer 41) used as a probe for the Pseudomonas *HCHL* gene during real time PCR analysis.

SEQ ID NO:81 is the nucleic acid sequence of the *ZmCesA10* promoter.

SEQ ID NO:82 is the nucleic acid sequence of the *ZmCesA11* promoter.

SEQ ID NO:83 is the nucleic acid sequence of the *ZmCesA12* promoter.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods and materials to produce para-hydroxybenzoic acid in the stalk tissue of genetically modified plants at commercially useful levels. Stem tissue-specific promoters have been identified from genes involved cellulose synthesis during plant secondary cell wall formation. Unexpectedly only promoters of certain cellulose synthase genes, when operably linked to an HCHL coding sequence, significantly limit HCHL expression to plant stem tissue. Promoter of genes controlling lignin biosynthesis in the plant stalk on the other hand failed to significantly increase stalk-specificity of HCHL expression. The use of cellulose synthase promoters for targeting HCHL expression to plant stem tissue resulted in significant pHBA production in the plants without the negative phenotypic changes associated with constitutive expression. A family of genes has been identified which represent a suitable source of stem tissue-specific promoters. Additionally, an HCHL enzyme from *Caulobacter crescentus* has been identified with superior catalytic efficiency for converting pHCACoA into pHBA.

The pHBA produced in the transgenic plants was converted to a mixture of pHBA glucoside (phenolic) and pHBA glucose ester by naturally occurring UDP-glucosyltransferases. Optionally, a foreign UDP-glucosyltransferase may be introduced into the transgenic plant for selective production of the pHBA glucose ester.

Transgenic plants (Arabidopsis) were modified to functionally express several chimeric genes encoding a 4-hydroxycinnamoyl-CoA hydratase/lyase

(HCHL). The chimeric genes were created by fusing various promoters to the coding sequence of the *HCHL* gene from *Pseudomonas putida* (DSM 12585). Several stem tissue-specific promoters were compared to constitutive promoters (non-tissue-specific) for their ability to 1) functionally express HCHL at levels comparable to the constitutive promoters for the production of pHBA, and 2) significantly limit expression of HCHL to plant stem tissue. The *Arabidopsis AtCesA7* (*IRX3*) promoter was shown to limit expression of HCHL to plant stem tissue. This parallels the expression pattern observed for the endogenous *AtCesA7* gene. Consequently, additional genes were identified as suitable sources of promoters for stem tissue-specific expression based on their observed expression patterns. Promoter sequences are provided that are suitable for driving tissue-specific HCHL expression. These include the Arabidopsis promoters derived from the *AtCesA4* (*IRX5*) and *AtCesA8* (*IRX1*) genes, as well as promoters from orthologous genes from maize and rice.

Methods are provided for the producing of pHBA from pHCACoA in plant stem tissue using an HCHL enzyme. Plant stem tissue is a natural reservoir where suitable levels of pHCACoA exist and where significant fluxes to the phenylpropanoid pathway can occur. Constitutive expression of HCHL (in all plant tissues) results in negative effects on the plant's agronomic performance. Methods are provided for tissue-specific expression of HCHL, resulting in production of pHBA in industrially-suitable amounts without negative phenotypic changes to the plant. Expression of HCHL needs to be limited to plant stem tissue. Tissues, such as leaf, do not contain suitable amounts of pHCACoA necessary for pHBA production. A unique set of tissue-specific promoters has been identified which are suitable for HCHL expression in plants.

The pHBA produced in the transgenic plants was converted to a mixture of pHBA glucoside (phenolic) and pHBA glucose ester by naturally occurring UDP-glucosyltransferases. Optionally, a foreign UDP-glucosyltransferase may be introduced into the transgenic plant for selective production of the pHBA glucose ester.

Definitions:

In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

"Polymerase chain reaction" is abbreviated "PCR".

"Para-hydroxybenzoic acid" or "p-hydroxybenzoic acid" is abbreviated "pHBA".

"Para-coumaroyl-CoA" is abbreviated "pHCACoA"

"Chorismate pyruvate lyase" is abbreviated "CPL" and refers to an enzyme which catalyzes the conversion of chorismate to pyruvate and pHBA.

"4-hydroxycinnamoyl-CoA hydratase/lyase" is abbreviated "HCHL" and refers to an enzyme (EC 4.2.1.101/EC 4.1.2.41) that catalyzes the hydration of the double bond of a hydroxycinnamoyl CoA thioester followed by a retro aldol cleavage reaction that produces a benzoyl aldehyde and acetyl CoA. The HCHL enzyme converts 1 mol of pHCACoA to 1 mol of acetylCoA and 1 mol of p-hydroxybenzaldehyde (pBALD). In plants, pBALD is subsequently converted to pHBA through the action of endogenous enzymes that are present in the cytoplasm.

"Homolog", "homologue", and "homologous gene" are terms used to describe a gene having similar structure, nucleic acid sequence, and evolutionary origin in comparison to another gene.

"Ortholog", "orthologue", and "orthologous gene" are terms used to describe a gene having similar structure, nucleic acid sequence, and evolutionary origin in comparison to another gene in a different species. Orthologs are homologs that usually share the same function and organization within a biosynthetic pathway. In the present invention, the orthologous genes encoding the subunits of the cellulose synthesis catalytic complex (associated with cells involved in the secondary cell wall synthesis) exhibit evolutionarily conserved structure, function, expression pattern, and organization. The conserved structure, function, expression pattern, and organization are believed to pre-date the evolutionary divergence of monocots and dicots. Promoters isolated from the *Arabidopsis thaliana* genes *AtCesA8 (IRX1), AtCesA7 (IRX3),* and *AtCesA4 (IRX5),* as well as promoters of the orthologous genes from maize and rice, are suitable for stem tissue-specific expression of HCHL.

"Paralog", "paralogue", and "paralogous gene" are terms used to describe a homolog where sequence divergence follows a gene duplication event within the same lineage. Paralogs are homologs that usually have different function.

"Cellulose synthase gene", "CESA", and "CesA" are terms used to describe a family of genes encoding proteins (EC 2.4.1.12) involved in cellulose synthesis. They generally exhibit significant homology to one another and share a conserved sequence motif (Taylor *et al.*, *supra* (2003)). The various members of this family (at least 12 identified in Arabidopsis) differ in their expression patterns and functions. Three CesA family members that encode for proteins involved in formation of the cellulose synthesis catalytic complex responsible cellulose production during secondary cell wall formation, have been identified in Arabidopsis (*AtCesA8*, *AtCesA7*, *AtCesA4*) as well as their orthologs from maize

and rice. *AtCesA8*, *AtCesA7*, and *AtCesA4* encode proteins that have been identified as absolutely necessary for cellulose synthesis in secondary cell wall formation. Expression of these three genes (as well as orthologs thereof) is significantly limited to cells involved in secondary cell wall biosynthesis (a significant portion of the cells in plant stem tissue). The promoters from these genes regulate an expression pattern suitable for recombinant HCHL expression in plant stem tissue.

"*AtCesA8*" and "*AtCesA8* (*IRX1*)" are terms used to describe one of the three genes identified in *Arabidopsis thaliana* encoding a cellulose synthase family protein that is a component of the cellulose synthesis catalytic complex. This gene, identified by Taylor *et al.* (*supra* (2003)) by an irregular xylem mutation "IRX1", is expressed in cells involved in secondary cell wall synthesis. The promoter from this gene exhibits a suitable tissue-specific expression pattern for driving recombinant HCHL expression in plant stem tissue.

"*AtCesA7*" and "*AtCesA7* (*IRX3*)" are terms used to describe one of the three genes identified in *Arabidopsis thaliana* encoding a cellulose synthase family protein that is a component of the cellulose synthesis catalytic complex. This gene, identified by Taylor *et al.* (*supra* (2003)) by an irregular xylem mutation "IRX3", is expressed in cells involved in secondary cell wall synthesis. The promoter from this gene exhibits a suitable tissue-specific expression pattern for driving recombinant HCHL expression in plant stem tissue.

"*AtCesA4*" and "*AtCesA4* (*IRX5*)" are terms used to describe one of the three genes identified in *Arabidopsis thaliana* encoding a cellulose synthase family protein that is a component of the cellulose synthesis catalytic complex. This gene, identified by Taylor *et al.* (*supra* (2003)) by an irregular xylem mutation "IRX5", is expressed in cells involved in secondary cell wall synthesis. The promoter from this gene exhibits a suitable tissue-specific expression pattern for driving recombinant HCHL expression in plant stem tissue.

"*ZmCesA10*" is a gene identified in *Zea mays* that is an ortholog of AtCesA4 (IRX5) based on comparative sequence analysis (Figure 4). The gene encodes a cellulose synthase family protein that is a component of the cellulose synthesis catalytic complex. *ZmCesA10* expression is limited to cells involved in synthesizing cellulose for secondary cell wall formation. The promoter from this gene exhibits a suitable tissue-specific expression pattern for driving recombinant HCHL expression in plant stem tissue.

"*ZmCesA11*" is a gene identified in *Zea mays* that is an ortholog of AtCesA8 (IRX1) based on comparative sequence analysis (Figure 4). The gene encodes a cellulose synthase family protein that is a component of the cellulose

synthesis catalytic complex. *ZmCesA11* expression is limited to cells involved in synthesizing cellulose for secondary cell wall formation. The promoter from this gene exhibits a suitable tissue-specific expression pattern for driving recombinant HCHL expression in plant stem tissue. "

"*ZmCesA12*" is a gene identified in *Zea mays* that is an ortholog of AtCesA7 (IRX3) based on comparative sequence analysis (Figure 4). The gene encodes a cellulose synthase family protein that is a component of the cellulose synthesis catalytic complex. *ZmCesA12* expression is limited to cells involved in synthesizing cellulose for secondary cell wall formation. The promoter from this gene exhibits a suitable tissue-specific expression pattern for driving recombinant HCHL expression in plant stem tissue.

"Rice orthologs" and "rice orthologous genes" are terms used to describe genes identified in *Oryza savita* (japonica cultivar group) which are orthologs to various maize cellulose synthase catalytic subunit genes (i.e. *ZmCesA10, Zm,CesA11,* and *ZmCesA12*) based on BLAST analysis of the publicly-available rice BAC database (National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine, Bethesda, MD). Based on the conserved nature of the expression patterns of the genes encoding proteins involved in the formation of the cellulose synthesis catalytic complex between monocots (*i.e. Zea mays*) and dicots (*i.e. Arabidopsis thaliana*) and the somewhat closer phylogenic relationship between maize and rice, the expression pattern of these genes are expected to parallel that of their orthologous counterparts in maize and Arabidopsis. The genes are believed to encode a cellulose synthase family protein that is a component of the cellulose synthesis catalytic complex. Suitable promoters derived from rice orthologs are those that significantly limit HCHL expression to plant stem tissue.

"Suitable tissue-specific promoter" is a term used in the present invention to describe a promoter that exhibits a stem tissue-specific expression pattern. Expression of chimeric genes created by the fusion of such a promoter to the *HCHL* coding sequence must be significantly limited to stem tissue cells where either suitable levels of pHCACoA exist or where large fluxes to the phenylpropanoid pathway can occur. Insignificant expression levels measured in non-stem tissues (especially leaf tissue) are acceptable as long as no detrimental effects on agronomic performance are observed. Preferred suitable tissue-specific promoters include those that exhibit the ability to preferentially expression active HCHL protein at least 10-fold higher in stem tissue in comparison to leaf tissue (stem:leaf $\geq$ 10:1). More preferred promoters are those that exhibit at least a 20-fold preference for HCHL expression in stem

tissue in comparison to leaf tissue (stem:leaf ≥ 20:1). Most preferred are those promoters that are capable of at least 50-fold stem to leaf tissue HCHL expression ratio (stem:leaf ≥ 50:1).

"*4CL1*" is the promoter from the gene encoding the 4CL enzyme. 4-Coumarate-coenzymeA ligase (4CL) enzymes are operationally soluble, monomeric enzymes of 60 kDa molecular weight belonging to the class of adenylate forming CoA ligases.

"*C4H*" is the promoter from the gene encoding the C4H enzyme. Cinnamate-4-hydroxylase (C4H) catalyzes the 4-hydroxylation of the aromatic ring of cinnamic acid.

"*C3'H*" is the promoter from the gene encoding the p-coumarate-3-hydrolyase enzyme. The p-coumarate-3-hydroxylase (C3H) enzyme (CYP98A3, GenBank® Accession No. AC011765) generates the 3,4-hydroxylated caffeoyl intermediate in lignin biosynthesis.

"*ACT2*" is a term used to describe the promoter from the *ACTIN2* gene. The promoter confers a constitutive pattern of reporter gene expression in plants (An *et al.*, *Plant Journal*, 10(1):107-121 (1996)).

"*35SCaMV*" is a term used to describe the promoter isolated from the Cauliflower Mosaic Virus that is commonly used in genetic engineering for constitutive expression of proteins.

"Cellulose synthesis catalytic complex" is a complex of at least 3 distinct cellulose synthase catalytic subunits that are required for secondary cell wall cellulose synthesis in plants. The genes encoding the members of this complex in Arabidopsis include *AtCesA4* (*IRX5*), *AtCesA7* (*IRX3*), *AtCesA8* (*IRX1*) (Taylor *et al.*, *supra* (2003)). All three subunits are required for correct assembly of the protein complex. The genes encoding the catalytic subunits exhibit a tissue-specific expression pattern suitable for HCHL expression. Corresponding orthologs in maize are shown by example to exhibit a similar expression pattern.

The terms "p-hydroxybenzoic acid glucoside" and "pHBA glucoside" refer to a conjugate comprising pHBA and a glucose molecule. pHBA glucose conjugates include the pHBA phenolic glucoside and pHBA glucose ester.

The terms "UDP-glucosyltransferase" and "glucosyltransferase" are abbreviated as "GT" and refer to enzymes (EC 2.4.1.194) involved in the formation of glucose-conjugated molecules. Such proteins catalyze a reaction between UDP-glucose and an acceptor molecule to form UDP and the glucosylated acceptor molecule. In most cases the hydroxyl group on C1 β-D-glucose is attached to the acceptor molecule via a 1-O-β-D linkage.

The term "aglycone" refers to substrates of the present invention that lack a glucose moiety (*i.e.*, unconjugated pHBA).

The term "pHBA derivative" refers to any conjugate of pHBA that may be formed in a plant as the result of the catalytic activity of the HCHL enzyme.

As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural, or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native" or "wild type" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" or "exogenous" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. "Foreign" may also be used to describe a nucleic acid sequence not found in the wild-type host into which it is introduced. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments that are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well-established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host.

Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites, and stem-loop structures.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence that can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters that cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". Tissue-specific promoters are those which direct expression of genes in limited tissue types. However, many "tissue-specific" promoters exhibit expression that is not significantly limited to the tissue of interest. Suitable tissue-specific promoters of the present invention are those that limit chimeric gene expression to stem tissue without significant expression in other tissues resulting in adverse phenotypic changes to the plant. The *Arabidopsis AtCesA8* (*IRX1*), *AtCesA7* (*IRX3*), and *AtCesA4* (*IRX5*) promoters, as well as promoters isolated from the respective orthologous genes from rice and maize (*ZmCesA11*, *ZmCesA12*, and *ZmCesA10*), are examples of suitable tissue-specific promoters useful in the present invention. The expression pattern associated with these promoters is highly correlated and significantly limited to plant stem tissue (Figure 5, Table 8). New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro

and Goldberg (In The Biochemistry of Plants, Vol. 15 published by Academic Press, Burlington, MA, pages 1-82, (1989)). It is further recognized that in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (*i.e.*, that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic", "recombinant", or "transformed" organisms.

The terms "plasmid", "vector", and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitates transformation of a particular host cell. "Expression

cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: 1.) Computational Molecular Biology (Lesk, A. M., Ed.) Oxford University: NY (1988); 2.) Biocomputing: Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY (1993); 3.) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., Eds.) Humana: NJ (1994); 4.) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and 5.) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI).

pHBA Production in Transgenic Plants using HCHL

pHBA is naturally occurring in nearly all plants, animals, and microorganisms, albeit in miniscule quantities. In plants, pHBA has been found in carrot tissue (Schnitzler *et al.*, *Planta*, 188:594, (1992)), in a variety of grasses and crop plants (Lydon *et al.*, *J. Agric. Food. Chem.*, 36:813, (1988)), in the lignin of poplar trees (Terashima *et al.*, *Phytochemistry*, 14:1991, (1972)), and in a number of other plant tissues (Billek *et al.*, *Oesterr. Chem.*, 67:401, (1966)). The fact that plants possess all of the necessary enzymatic machinery to synthesize pHBA suggests that they may be a useful platform for producing this monomer. For example, as a renewable resource a plant platform would require far less energy and raw materials than either petrochemical or microbial methods for producing the monomer. Similarly, a plant platform represents a far greater available biomass for monomer production than a microbial system. Finally, the natural presence of pHBA in plants suggests that host toxicity (a result of overproduction of the compound) might not be a problem.

Transgenic plants that accumulate significantly higher levels of pHBA than wild-type plants have been described. 4-Hydroxycinnamoyl-CoA

hydratase/lyase (HCHL) isolated from *Pseudomonas fluorescens* AN103 is a bacterial enzyme that when expressed in transgenic tobacco (*Nicotiana tabacum* cv. Xanthi XHFD8) resulted in significant accumulation of pHBA glucosides (Mayer *et al.*, *supra*). Expression of HCHL in the transgenic plant's cytosol redirected the carbon flux from the phenylpropanoid pathway into the production of pHBA glucosides. However, constitutive expression of HCHL in plant tissues (such as leaf) where inadequate amounts of pHCACoA exist or where a high-flux to the phenylpropanoid pathway cannot occur, significantly depletes of secondary metabolites having roles as plant growth regulators, UV protectants, or cell wall components such as lignin, cutin, or suberin. Depletion of secondary metabolites in these tissues resulted in adverse plant growth defects such as interveinal leaf chlorosis, stunting, low pollen production, and male sterility.

Sterility is very likely caused by severe reduction in flavonoid levels. For example, pHCACoA-derived flavonols are required for pollen germination in solanaceous plants like tobacco (Napoli *et al.*, *Plant Physiology*, 120(2):615-622 (1999)). Premature senescence and dwarfism may be caused by the depletion of ferulic acid-derived dehydrodiconiferyl alcohol glucosides (Teutonico *et al.*, P*lant Physiology*, 97(1):288-97 (1991)). There is evidence that these molecules are components of a cytokinin-mediated regulatory circuit controlling cell division in plants (Teutonico *et al.*, *supra*). (Cytokinin is obviously an important signaling component that counteracts senescence (Gan and Amasino, *BioEssays*, 18(7):557-565 (1996))). The cytokinin-like activity of these molecules could lead one to speculate that their depletion is also responsible for the early-senescence phenotype of some HCHL-expressing plants.

The source of the *HCHL* gene used for engineering transgenic plants for pHBA production is not limited to *Pseudomonas fluorescens* AN103 (Gasson *et al.*, *J Bio Chem*, 273(7):4163-4170 (1998)); WO 97/35999; and US 6323011). Additional microorganisms reported to have genes encoding HCHL activity include, but are not limited to, *Pseudomonas putida* DSM 12585 (Muheim and Lerch, *Appl Microbiol Biotechnol*, 51:456-461 (1999)), *Pseudomonas putida* WCS358 (Venturi *et al.*, *Microbiol*, 144(4):965-973 (1998)); *Pseudomonas* sp. HR199 (Priefert *et al.*, *J Bacteriol*, 179(8):2595-2607 (1997)), *Delftia acidovorans* (Plagenborg *et al.*, *FEMS Microbiol Lett*, 205(1):9-16 (2001)), and *Amycolatopsis* HR167 (Achterholt *et al.*, *Appl Microbiol Biotechnol*, 54(6): 799-807 (2000); WO 01/044480).

The use of the *HCHL* gene from *Pseudomonas putida* DSM 12585 is illustrated in the present invention. However, the source of suitable *HCHL* genes useful for plant transformation and production of pHBA is not limited to

the examples provided herein. Examples include, but are not limited to, those *HCHL* genes listed in Table#1. The coding sequence from any *HCHL* gene is suitable in the present invention based on the reported ability to functionally express various bacterial *HCHL* genes in the cytosol of plant cells (Mitra *et al., supra*; Mayer *et al., supra*; and WO 97/35999). Additionally, an HCHL isolated from *Caulobacter crescentus* (SEQ ID NOs:60 and 61) is provided that exhibits increased kinetic properties for pHBA synthesis as compared to the HCHL enzymes from *P. putida* DSM 12685 and *P. fluorescens* AN103.

Table 1
Source of *HCHL* Genes

GenBank® Accession Number and (Source Organism)	Sequence Identification Number (SEQ ID NO)
(*Pseudomonas putida* DSM 12585)	5
Y13067 (*Pseudomonas fluorescens* AN103)	58
Y14772 (*Pseudomonas putida* WCS358)	59
AE005909.1 (*Caulobacter crescentus*)	60
Y11520.1 (*Pseudomonas* sp. HR199)	62
AJ300832 (*Delftia acidovorans*)	63
AJ290449 (*Amycolatopsis* sp. HR167)	64

HCHL Expression Cassette

An expression cassette useful for the producing of pHBA in plant stem tissue includes a suitable stem tissue-specific promoter operably linked to the *HCHL* coding sequence. Typically, the expression cassette will comprise (1) the cloned *HCHL* coding sequence under the transcriptional control of 5' (suitable stem cell specific promoter) and 3' regulatory sequences and (2) a dominant selectable marker. The present expression cassette may also contain a transcription initiation start site, a ribosome-binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Optionally, the cassette may also comprise one or more introns in order to facilitate HCHL expression.

The most well characterized HCHL gene has been isolated from *Pseudomonas fluorescens* AN103 (GenBank® Accession No. Y13067.1). DNA sequence of an HCHL gene from *Pseudomonas putida* DSM 12585 (Muheim and Lerch, *App. Micro. Biotech.*, 51(4):456-461 (1999)) and the deduced amino acid sequence of the HCHL protein of this organism is set forth herein as SEQ ID NO:5 and SEQ ID NO:6, respectively. This gene has been isolated by the Applicants and is useful for producing of pHBA in transgenic plants.

Tissue-specific Promoters for Expression of *HCHL*

The use of tissue-specific promoters is known in the art. However, many of these reported promoters exhibit only preferential expression in certain plant and/or animal tissues, allowing significant expression in other tissues, albeit at levels at or below the target tissue. HCHL expression in this invention is selectively limited to where a suitable substrate pool is available or where large fluxes to the phenylpropanoid pathway may occur since expression in other tissues, such as leaf, has been shown to be detrimental to the agronomic performance of the plant (Mayer *et al.*, *Plant Cell*, 13:1669-1682 (2001)).

Genes involved in lignin biosynthesis were tested as a source of suitable tissue-specific promoters. These promoters were operably linked to the coding sequence of an *HCHL* gene. The chimeric constructs were tested for tissue-specific expression in plants (Arabidopsis). HCHL expression was not significantly limited to plant stem tissue (Table 6). Because of this, these promoters were not considered suitable for HCHL expression.

Plant stem tissue contains significant amount of cellulose. Genes encoding enzyme involved in cellulose synthesis were identified as possible source for tissue-specific promoters suitable for chimeric HCHL expression. Three genes from *Arabidopsis thaliana* were identified as critical to cellulose synthesis in cells involved in secondary cell wall formation. These genes, *AtCesA4* (*IRX5*), *AtCesA7* (*IRX3*), and *AtCesA8* (*IRX1*), have been shown to have a desirable expression pattern suitable for chimeric *HCHL* expression. The proteins encoded by these genes interact and form the cellulose synthesis catalytic complex (Taylor *et al.*, *supra* (2003)). Their expression is closely correlated with one another, essentially limited to cells involved in producing cellulose for secondary cell wall formation. The promoter from one of the genes (*AtCesA7* (*IRX3*)), was isolated and operably linked to the coding sequence of an *HCHL* gene. HCHL expression was measured between the stem tissue

versus leaf tissue in Arabidopsis plants transformed with the chimeric constructs. HCHL expression was not significantly limited to plant stem tissue (Table 6).

It has been reported that the genetic organization of the genes encoding proteins involved in forming the cellulose synthesis catalytic complex have been conserved across monocots and dicots (Holland *et al.*, *Plant Physiol.*, 123:1313-1323 (2000)). Expression analysis of orthologous genes from *Zea mays* (*ZmCesA10*, *ZmCesA11*, and *ZmCesA12*) also shows a similar pattern, namely gene expression that is essentially limited to cells involved in cellulose synthesis during secondary cell wall formation. Consequently, promoters from these genes as well as promoters from orthologous genes from rice (*Oryza savita* (japonica cultivar group)) are suitable for stem-specific expression of HCHL.

In the present invention, suitable promoters for HCHL expression control an HCHL expression pattern - where HCHL activity is at least 20-fold higher in stem tissue in comparison to leaf tissue. More preferred promoters are those that control an HCHL expression pattern where HCHL activity is at least 30-fold higher in stem tissue when compared to leaf tissue. Most preferred promoters suitable in the present invention are those that control an HCHL expression pattern where HCHL activity is at least 50-fold higher in stem tissue when compared to leaf tissue. Suitable promoters can be identified by comparison of HCHL activity converting pHCACoA to p-Hydroxybenzaldehyde (expressed as pkcat/mg protein) in stem and leaf tissue of transgenic plants expressing HCHL genes under the control of tissue-specific promoters. Alternatively, suitable promoters can be identified by comparing pHBA production observed in stem and leaf tissues of transgenic plants expressing HCHL genes under the control of tissue-specific promoters. Suitable promoters when fused to HCHL genes will generate a pattern of pHBA accumulation where pHBA accumulation in stalk tissue is > 10 higher than pHBA accumulation in leaf tissue. Alternatively, suitable promoters can be identified by performing MPSS analysis of gene expression in various plant tissues. Promoters suitable for HCHL gene expression in plants are those that show high levels of gene expression in stalk tissue (> 350 ppm) and show a pattern of gene expression were gene expression is at least 10-fold higher in stalk tissue when stalk and leaf tissues are compared. More preferred promoters show a pattern of gene expression were gene expression is at least 20-fold higher in stalk tissue when stalk and leaf tissues are compared. More preferred promoters show a pattern of gene expression were gene expression is at least 50-fold higher in stalk tissue when stalk and leaf tissues are compared.

UDP-Glucosyltransferases

Most of the products of secondary metabolism in plants are glycosylated (Harborne, J., *Introduction to Ecological Biochemistry*, 4th ed.; Academic Press: London, 1993), as are many herbicides after modification by phase I enzymes. An impressive array of conjugated species, including coumaryl glucosides, flavonoids, anthocyanins, cardenolides, soponins, cyanogenic glucosides, glucosinolates, and betalains, are known to be stored in the vacuole (Wink, M., In *The Plant Vacuole: Advances in Botanical Research*; Leigh, R. A., Sanders, D. and Callow, J. A., Eds.; Academic Press: London, New York, 1997; Vol. 25, pp 141-169). Based on these observations and the fact that most UDP-glucosyltransferases are located in the cytosol, glucosylation has been invoked as a prerequisite for uptake and accumulation in the vacuole. In addition, *in vitro* experiments clearly demonstrate that isolated vacuoles and/or vacuolar membrane vesicles are able to take up certain glucose conjugates, while the parent molecules are not transported (Wink, M., *supra*).

It has been shown that the vast majority of pHBA produced in transgenic plant cells is rapidly converted by endogenous UDP-glucosyltransferases to two glucose conjugates, a phenolic glucoside with the glucose moiety attached to the aromatic hydroxyl group, and a glucose ester where the sugar is attached to the aromatic carboxyl group (Siebert *et al.*, *Plant Physiol.* 112:811-819 (1996); Mayer *et al.*, *supra*; Mitra *et al.*, *supra*; and US SN 10/359369). The vast majority of plants contain endogenous UDP-glucosyltransferases that form both glucose conjugates of pHBA. Although both glucose conjugates accumulate in the vacuole, they have very different chemical properties and physiological roles.

For example, the pHBA glucose ester (like other acetal esters) is characterized by high free energy of hydrolysis, which makes it very simple to recover the parent compound with low concentrations of either acid or base. This could greatly reduce the cost of producing pHBA in plants. Furthermore, it is well established that certain glucose esters are able to serve as activated acyl donors in enzyme-mediated transesterification reactions (Li *et al.*, *Proc. Natl. Acad. U.S.A.*, 97(12):6902-6907 (2000); Lehfeldt *et al.*, *Plant Cell*, 12(8):1295-1306 (2000)). In light of these observations, it would be extremely desirable to control the partitioning of pHBA glucose conjugates *in vivo*. For example, by overexpressing an appropriate glucosyltransferase in transgenic plants that generate large amounts of pHBA, it might be possible to accumulate all of the desired compound as the glucose ester, which can be easily hydrolyzed to free pHBA. While the above scenario is extremely attractive, it

requires an enzyme with the appropriate properties and molecular information that would allow access to the gene (*e.g.*, its nucleotide or primary amino acid sequence).

Commonly owned US SN 10/359369, hereby incorporated by reference, provides examples of UDP-glucosyltransferases that preferentially use pHBA as a substrate and which selectively convert pHBA to pHBA glucose ester. Examples of nucleic acid molecules encoding these pHBA UDP-glucosyltransferases are represented by SEQ ID NOs:65, 66, and 67, respectively. In a preferred embodiment of the invention, genes encoding pHBA UDP-glucosyltransferases that preferentially convert pHBA to pHBA glucose ester are used to transform plants functionally expressing HCHL in plant stem tissue.

Plant Gene Expression

Promoters useful for expressing the genes are numerous and well known in the art. Plant tissue-specific promoters have been reported (Yamamoto *et al.*, *Plant Cell Phys.* 35(5):773-778 (1994); Kawamata *et al.*, *Plant Cell Phys.*, 38(7):792-803 (1997); Rinehart *et al.*, *Plant Phys.*, 112:1331-1341 (1996); Van Camp *et al.*, *Plant Phys.*, 112:525-535 (1996); Canevascini *et al.*, *Plant Phys.*, 112:513-524 (1996); Guevara-Garcia *et al.*, *Plant Journal*, 4(3):495-505 (1993); and Yamamoto *et al.*, *Plant Journal*, 12(2):255-265 (1997)). However, the ability of these promoters to limit HCHL expression to plant stem tissue has not been reported. It has been shown that HCHL expression must be limited to plant tissues where a significant pool of substrate (pHCACoA) is available and where high flux to the phenylpropanoid pathway is possible.

A preferred embodiment of the current invention is the use of an exogenous UDP-glucosyltransferase for selection production of pHBA glucose ester (US SN 10/359369). Any combination of any promoter and any terminator capable of inducing expression of the exogenous UDP-glucosyltransferase may be used in the present invention. Expression of an exogenous pHBA UDP-glucosyltransferase does not need to be targeted to a specific plant tissue. Some suitable examples of promoters and terminators include those from nopaline synthase (*nos*), octopine synthase (*ocs*), and cauliflower mosaic virus (*CaMV*) genes. Such promoters, in operable linkage with the pHBA UDP-glucosyltransferases of the present invention, are capable of promoting expression of these genes for selective production of pHBA glucose ester. High-level plant promoters that may be also be used in this invention include the promoter of the small subunit (ss) of the ribulose-1,5-bisphosphate carboxylase from soybean (Berry-Lowe *et al.*, *J. Mol. App. Gen.*, 1:483-498 (1982)), and the

promoter of the chlorophyll a/b binding protein. These two promoters are known to be light-induced in plant cells (See, for example, Genetic Engineering of Plants, an Agricultural Perspective, A. Cashmore, Plenum, New York (1983), pages 29-38; Coruzzi, G. *et al.*, *J. Bio. Chem.*, 258:1399 (1983); and Dunsmuir *et al.*, *J. Mol. App. Gen.*, 2:285 (1983)).

Where polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of each gene's coding region in the present invention. The polyadenylation region can be derived from a variety of plant genes or from T-DNA. For example, the 3' end sequence to be added can be derived from the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or, less preferably, from any other eukaryotic gene.

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inducing a spliceable intron in the transcription unit of both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg, *Mol. Cell Biol.*, 8:4395-4405 (1988); Callis *et al.*, *Genes Dev.*, 1:1183-1200 (1987)). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. (See generally, The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994).)

Virtually any plant host that is capable of supporting the expression of the genes in the present invention will be suitable; however, crop plants are preferred for their ease of harvesting and large biomass. Suitable plant hosts include, but are not limited to, both monocots and dicots such as soybean, rapeseed (*Brassica napus*, *B. campestris*), sunflower (*Helianthus annus*), cotton (*Gossypium hirsutum*), corn, tobacco (*Nicotiana tabacum*), alfalfa (*Medicago sativa*), wheat (*Triticum sp*), barley (*Hordeum vulgare*), oats (*Avena sativa, L*), sorghum (*Sorghum bicolor*), rice (*Oryza sativa*), *Arabidopsis,* sugar beet, sugar cane, canola, millet, beans, peas, rye, flax, and forage grasses. Preferred plant hosts are tobacco, *Arabidopsis thaliana*, sugarcane, and sugar beet.

Plant Transformation

A variety of techniques are available and known to those skilled in the art to introduce constructs into a plant cell host. These techniques include transformation with DNA employing *A.tumefaciens* or *A. rhizogenes* as the transforming agent, electroporation, and particle acceleration (EP 295959 and

EP 138341). One suitable method involves the use of binary type vectors of Ti and Ri plasmids of *Agrobacterium sp*. Ti-derived vectors transform a wide variety of higher plants including monocotyledonous and dicotyledonous plants such as soybean, cotton, rape, tobacco, and rice (Pacciotti *et al.*, *Bio/Technology*, 3:241 (1985); Byrne *et al.*, *Plant Cell, Tissue and Organ Culture*, 8:3 (1987); Sukhapinda *et al.*, *Plant Mol. Biol.*, 8:209-216 (1987); Lorz *et al.*, *Mol. Gen. Genet.*, 199:178 (1985); Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183 (1985); Park *et al.*, *J. Plant Biol.*, 38(4):365-71 (1995); and Hiei *et al.*, *Plant J.*, 6:271-282 (1994)). The use of T-DNA to transform plant cells has received extensive study and is amply described (EP 120516; Hoekema, In: The Binary Plant Vector System, Offset-drukkerij Kanters B.V.; Alblasserdam (1985), Chapter V; Knauf *et al.*, *Genetic Analysis of Host Range Expression by Agrobacterium* In: Molecular Genetics of the Bacteria-Plant Interaction, Puhler, A. ed., Springer-Verlag, New York, 1983, p. 245; and An *et al.*, *EMBO J.*, 4:277-284 (1985)). For introduction into plants, the chimeric genes can be inserted into binary vectors as described in the examples.

Other transformation methods are known to those skilled in the art. Examples include direct uptake of foreign DNA constructs (EP 295959), techniques of electroporation (Fromm *et al.*, *Nature* (London), 319:791 (1986)), and high-velocity ballistic bombardment with metal particles coated with the nucleic acid constructs (Kline *et al.*, *Nature* (London), 327:70 (1987); and US 4945050). Once transformed, the cells can be regenerated by those skilled in the art. Of particular relevance are the recently described methods to transform foreign genes into commercially important crops, such as rapeseed (De Block *et al.*, *Plant Physiol.*, 91:694-701 (1989)), sunflower (Everett *et al.*, *Bio/Technology*, 5:1201 (1987)), soybean (McCabe *et al.*, *Bio/Technology*, 6:923 (1988); Hinchee *et al.*, *Bio/Technology* 6:915 (1988); Chee *et al.*, *Plant Physiol.*, 91:1212-1218 (1989); Christou *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:7500-7504 (1989); EP 301749), rice (Hiei *et al.*, *supra*), and corn (Gordon-Kamm *et al.*, *Plant Cell,* 2:603-618 (1990); and Fromm *et al.*, *Biotechnology,* 8:833-839 (1990)).

Transgenic plant cells are placed in an appropriate medium to select for the transgenic cells that are then grown to callus. Shoots are grown from callus and plantlets generated from the shoot by growing in rooting medium. The various constructs normally will be joined to a marker for selection in plant cells. Conveniently, the marker may be resistance to a biocide (particularly an antibiotic such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, herbicide, or the like). The particular marker used will select for transformed

cells as compared to cells lacking the DNA that has been introduced. Components of DNA constructs including transcription cassettes may be prepared from sequences which are native (endogenous) or foreign (exogenous) to the host. Heterologous constructs will contain at least one region that is not native to the gene from which the transcription-initiation-region is derived. To confirm the presence of the transgenes in transgenic cells and plants, a Southern blot analysis can be performed using methods known to those skilled in the art.

Promoters from Orthologs of *Arabidopsis AtCesA8 (IRX1), AtCesA7 (IRX3), and AtCesA4 (IRX5)* Genes

The proteins (catalytic subunits) involved in forming the cellulose synthesis catalytic complex are encoded by three genes (Taylor *et al.*, *supra* (2003)). In *Arabidopsis thaliana* these genes have been designated *AtCesA8* (*IRX1*), *AtCesA7* (*IRX3*), and *AtCesA4* (*IRX5*) using the current naming convention ("*At*" = *Arabidopsis thaliana*; "*CesA*" = cellulose synthase gene followed by an assigned number designation; Delmer, DP., *Annu Rev Plant Physiol Plant Mol Biol.*, 50:245-276 (1999)). The roles these genes play in cellulose biosynthesis in secondary cell wall formation were identified by the mutations effecting xylem formation (irregular xylem; *IRX1*, *IRX3*, and *IRX5*, corresponding to *AtCesA8*, *AtCesA7*, and *AtCesA4*; respectively) (Taylor *et al.*, *supra* (2003); Taylor *et al.*, *supra* (2000); and Richmond and Somerville, *supra*). The expression pattern comparisons of these genes, and corresponding orthologs in other plants, indicates that 1) there is a high correlation between the expression of these genes and the tissue in which they are expressed and 2) their expression is essentially limited to stem tissue in both monocots and dicots. In *Arabidopsis* (dicot), Taylor *et al.* (*supra* (2003)) illustrate how *AtCesA8* (*IRX1*), *AtCesA7* (*IRX3*), and *AtCesA4* (*IRX5*) expression is essentially limited to stem tissue. Orthologs from maize (monocot), namely *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* exhibit the same expression pattern, indicating that the functional relationship and tissue-specificity has been evolutionarily conserved (Example 5; Figure 4). Groupings of *CesA* orthologs show greater similarity than paralogs (Holland *et al.*, *supra*). As shown in Figure 4, both monocots and dicots group within the same classes when comparing plant cellulose synthase proteins, indicating that the divergence into at least some of these subclasses may have arisen relatively early in the evolution of these genes (Holland *et al.*, *supra*).

Rice (*Oryza sativa* (japonica cultivar group)) has orthologs of the maize *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* genes. Based on the conserved expression patterns observed between *Arabidopsis* and maize and the

somewhat closer phylogenic relatedness between maize and rice (both monocots), promoters from orthologous rice genes were identified by sequence analysis using the maize *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* genes. A comparison of the respective gene from *Arabidopsis*, maize, and rice is provided in Table 2. The promoter sequences for the *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* genes were identified by sequencing genomic DNA upstream of the start codon for each respective gene. The promoter sequences for the *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* promoters are provided as SEQ ID NOs:81, 82, and 83, respectively. The respective rice promoter sequences (defined in the present invention as the 2500 bp 5' to the start codon of each respective ortholog) are provided as SEQ ID NOs:43, 44, and 45.

Table 2. Orthologous Genes from *Arabidopsis*, Maize (*Zea mays*), and Rice (*Oryza savita*) Associated with the Formation of the Cellulose Synthesis Catalytic Complex

	Corresponding Orthologs Identified from:	
Arabidopsis thaliana Gene	*Zea mays*	*Oryza savita*
AtCesA8 (*IRX1*)	*ZmCesA11* (SEQ ID NO:33)	Rice ortholog of *ZmCesA11* (SEQ ID NO:39)
AtCesA7 (*IRX3*)	*ZmCesA12* (SEQ ID NO:35)	Rice ortholog of *ZmCesA12* (SEQ ID NO:41)
AtCesA4 (*IRX5*)	*ZmCesA10* (SEQ ID NO:31)	Rice ortholog of *ZmCesA10* (SEQ ID NO:37)

Gene Expression Analysis

Gene expression analysis of various cellulose synthase genes has been reported. Taylor *et al.* (*PNAS*, 100(3):1450-1455 (2003) and *Plant Cell*, 12:2529-2539 (2000)) reported that proteins encoded by the *Arabidopsis* cellulose synthase genes encoding proteins forming the cellulose synthesis catalytic complex (*AtCesA8*, *AtCesA7*, and *AtCesA4*) are co-expressed in exactly the same cells. The data indicates that the promoters from these genes are suitable for stem tissue expression.

Orthologs from maize exhibit a nearly identical tissue-specific expression pattern in comparison to *Arabidopsis* (Figures 4 and 5; Table 9) as illustrated by MPSS (Lynx Therapeutics, Hayward, CA) analysis (Brenner *et al.*, *PNAS*, 97(4):1665-1670 (2000); US 6,265,163; and US 6,511,802; hereby incorporated by reference). MPSS is a technique in which cDNA is attached to the surface of

a unique microbead. Highly expressed mRNA is represented on a proportionally larger number of microbeads. Signature sequences of approximately 16-20 nucleotides are then obtained from these microbeads by iterative cycles of restriction with a type IIs endonuclease, adaptor ligation, and hybridization with encoded probes. cDNA was collected from various maize tissues and analyzed by MPSS. The level of expression of a gene is determined by the abundance of its signature in the total pool (Figure 5, Example 5, Table 9).

The expression levels of active HCHL from the genetic construct comprising the *AtCesA7* (*IRX3*) promoter operably linked to an *HCHL* coding sequence were indirectly measured by comparing the enzymatic activity of the expressed HCHL protein isolated from stem and leaf tissue from a transformed model plant (Table 6).

Promoters derived from the *AtCesA8*, *AtCesA7*, and *AtCesA4* genes and the promoters derived from the corresponding orthologous genes from maize and rice exhibit suitable tissue-specific expression patterns useful for stem tissue-specific HCHL expression.

Enzyme Kinetics

Important parameters of enzyme-catalyzed reactions include 1) turnover number (Kcat), a unit for catalytic power of a monomeric enzymatic catalyst expressed as µmol of product formed per second per µmol of enzyme, and 2) Km, a unit for affinity of the enzyme to a particular substrate, expressed as the substrate concentration at which 50 % of maximum velocity is achieved. Catalytic efficiency is usually expressed as Kcat/Km. The greater the value of Kcat/Km, the more rapidly and efficiently the substrate is converted into product.

Expression of Divergent HCHL Sequences

Cosuppression suppression, also known as sense suppression, is a phenomenon that can occur at the transcriptional or post-transcriptional level. One major factor that determines whether or not post-transcriptional silencing occurs is the level of homology between coding sequences of homologous genes. Decreasing the level of sequence homology between coexpressed genes correlates with a decrease in post-transcriptional gene silencing. Thierry and Vaucheret (*Plant Mol. Biol.*, 32:1075-1083) describe how post-transcriptional gene silencing was observed when two genes sharing 84% identity were coexpressed while a transgene sharing only 76% identity to an endogenous plant gene escaped cosuppression. Niebel *et al.* (*Plant Cell*, 7:347-358 (1995)) described how selective cosuppression may occur as a consequence of the higher degree of DNA sequence identity. Genes having

coding sequences sharing 81% identity were cosuppressed while those sharing 63% identity were not.

Applicants disclose (in Example 2) that HCHL expression level and not abundance of the HCHL substrate pHCACoA limits pHBA accumulation in the plant stalk. Thus, further improvements in pHBA accumulation in the plant stalk could be achieved by introducing of DNA elements that consist of multiple HCHL expression cassettes each comprised of suitable promoter, an HCHL coding sequence, and a terminator sequence. Promoters and HCHL coding sequences in the expression cassettes need to be divergent in sequence in order to avoid transcriptional and post-transcriptional gene silencing effects that are triggered when identical or highly similar genes are expressed in the same eukaryotic cell. Applicants provide both divergent promoters (of cellulose synthase genes) and an HCHL gene of *Caulobacter crescentus* that shares only 57 % sequence identity to HCHL genes from Pseudomonas. Applicants predict that DNA elements containing two different HCHL genes from Pseudomonas and Caulobacter under the control of different cellulose synthase promoters would provide a route to pHBA accumulation in the plant stalk that would exceed that observed with DNA elements containing only one HCHL gene or two closely related HCHL genes.

Description of Preferred Embodiments

Examples 1 and 2 illustrate the isolation and effects of constitutive expression of an HCHL gene from *Pseudomonas putida* (DSM 12585) on plant development. Enzymatic activity and pHBA accumulation are compared to show that HCHL is substrate-limited in plant leaf tissue, confirming the observation that constitutive HCHL expression produces negative phenotypic changes to the plant.

Example 3 provides a comparison of several tissue-specific promoters. Of the various HCHL expression cassettes assayed, only the chimeric gene comprising a promoter isolated from the *Arabidopsis thaliana AtCesA7* (*IRX3*) gene exhibited suitable tissue-specific expression. The *AtCesA7* (*IRX3*) gene has been reported to exhibit a suitable tissue-specific expression pattern, identical to the desired expression pattern for stem-specific expression of HCHL. Two additional genes isolated from *Arabidopsis thaliana*, namely *AtCesA4* (*IRX5*) and *AtCesA8 (IRX1*), have been reported to have nearly identical expression to that of *AtCesA7* (*IRX3*) (Taylor *et al.*, *supra* (2003)). These three genes encode cellulose synthesis catalytic subunits. Expression of these genes is normally limited to cells involved in plant secondary cell wall formation in the

vascular tissue (stem tissue). Promoters from these genes were identified as suitable for tissue-specific HCHL expression.

Orthologous genes exhibiting a conserved expression pattern, sequence similarity, and function were identified in *Zea mays* (Examples 4 and 5; Figure 4). Phylogenic analysis revealed that the structure, function, and overall organization in the cellulose synthesis pathway were evolutionarily conserved suggesting that this conserved relationship predates that divergence of monocots and dicots. The promoters from *Zea mays* genes *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* are suitable for creating chimeric HCHL expression cassettes.

Examples 6 and 7 illustrate the identification of orthologous rice genes that are expected to have similar structure, function, and overall organization in the cellulose synthesis pathway in comparison to genes from *Zea mays*. Closely related genes were identified which are orthologs of the *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* genes. The promoters were identified as those sequences approximately 2500 bp 5' to the gene's coding sequence.

Prophetic Example 8 provides a method to create various chimeric HCHL constructs using the suitable tissue-specific promoters identified previously. This method is an example of how to create suitable HCHL expression cassettes. One skilled in the art can easily recognize that the source of HCHL gene is not limited to that which is provided in the examples *(i.e., Pseudomonas putida* DSM 12585).

pHBA by HCHL in stalk tissue is limited by enzymatic activity, even when stalk-specific promoters are used. Example 9 provides comparative enzyme kinetic data for HCHL enzymes from *Pseudomonas putida* (DSM 12585), *Pseudomonas fluorescens* AN103, and *Caulobacter crescentus* (previously uncharacterized). Kinetic analysis revealed that the HCHL from *C. crescentus* has superior catalytic efficiency (Kcat/Km) when compared to the other enzyme sources (50% improvement).

The present methods illustrate the creation of an HCHL expression cassette: the expression cassette comprising a tissue-specific promoter operably linked to an HCHL coding sequence. Numerous sources of suitable HCHL genes are known in the art. Several examples are provided in Table 1. Preferred are HCHL genes isolated from a bacterium selected from the group consisting of Pseudomonas, Caulobacter, Delftia, Amycolatopsis, and Sphingomonas. More preferred sources of HCHL genes are *Pseudomonas putida* (DMS 12585), *Pseudomonas fluorescens* AN103, *Pseudomonas putida* WCS358, *Caulobacter crescentus*, Pseudomonas sp. HR199, *Delftia*

acidovorans, *Amycolatopsis* sp. HR167, and *Sphingomonas paucimobilis*. Most preferred sources of HCHL genes are *Pseudomonas putida* (DSM 12585) and *Caulobacter crescentus*.

The tissue-specific promoters of the present invention useful for expressing an HCHL enzyme in plant stem tissue are those isolated from genes encoding a subunit of the cellulose synthesis catalytic complex involved in the synthesis of cellulose during plant secondary cell wall formation in the plant vascular tissue (stem tissue). Preferred tissue-specific promoters are isolated from *Arabidopsis thaliana* genes *AtCesA4*, *AtCesA7*, and *AtCesA8*; *Zea mays* genes *ZmCesA10*, *ZmCesA11*, and *ZmCesA12*; and the *Oryza savita* orthologs of *ZmCesA10*, *ZmCesA11*, and *ZmCesA12*. More preferred tissue-specific promoters are isolated from *AtCesA4*, *AtCesA7*, *AtCesA8*, the *Oryza savita* ortholog of *ZmCesA10*, the *Oryza savita* ortholog of *ZmCesA11*, and the *Oryza savita* ortholog of *ZmCesA12*. Even more preferred are the promoters isolated from the *AtCesA4*, *AtCesA7*, and *AtCesA8*. Most preferred is the promoter isolated from *AtCesA7*.

Plant suitable for production of pHBA using the present methods include tobacco, *Arabidopsis*, sugar beet, sugar cane, soybean, rapeseed, sunflower, cotton, corn, alfalfa, wheat, barley, oats, sorghum, rice, canola, millet, beans, peas, rye, flax, and forage grasses. Preferred plant hosts are tobacco, *Arabidopsis thaliana*, sugar cane, and sugar beet.

The pHBA produced within the plant is rapidly glucosylated by a pHBA UDP-glucosyltransferase into the pHBA glucoside or pHBA glucose ester for storage in the plant's vacuoles. The UDP-glucosyltransferase can be either endogenous or foreign to the plant. Preferred are recombinant UDP-glucosyltransferases that preferentially catalyze the formation of pHBA glucose ester. More preferred are those recombinant UDP-glucosyltransferase gene isolated from *Vitis* sp., Eucalyptus grandis, and *Citrus mitis*. More preferred are those UDP-glucosyltransferases represented by SEQ ID NOs:65, 66, and 67. The pHBA glucose ester can be easily hydrolized to form unconjugated pHBA. Expression of a recombinant pHBA UDP-glucosyltransferase is not limited to the use of stem specific promoters.

Lastly, the low level (<57%) sequence identity of the HCHL coding sequences of *Pseudomonas putida* (DSM 12585) and Pseudomonas fluorescens AN103 relative to the HCHL coding sequence of *Caulobacter crescentus* is expected to allow co-expression of both HCHL genes (i.e. without sense suppression) in the same plant providing an additional means to increase pHBA production in plant stem tissue. Preferably, the HCHL genes targeted for

coexpression should have less than 70% sequence identity between coding sequences. More preferably, the sequence identity should be less than 65%. Most preferably, the sequence identity is less than 60%.

GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989) (hereinafter "Maniatis"); and by Silhavy, T. J., Bennan, M. L., and Enquist, L. W., Experiments with Gene Fusions, Cold Spring Harbor Laboratory Cold Press Spring Harbor, NY (1984) (hereinafter "Silhavy"); and by Ausubel, F. M. *et al.*, Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience (1987) (hereinafter "Ausubel").

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg, and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC (1994)) or Brock (*supra*). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO), unless otherwise specified.

Manipulations of genetic sequences were accomplished using the suite of programs available from the Genetics Computer Group Inc. (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI). The GCG program "Pileup" used the gap creation default value of 12, and the gap extension default value of 4. The CGC "Gap" or "Bestfit" programs used the default gap creation penalty of 50 and the default gap extension penalty of 3. In any case where GCG program parameters were not prompted for, in these or any other GCG program, default values were used.

Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences is performed using the Clustal method of alignment (Higgins and Sharp, *CABIOS.*, 5:151-153 (1989); Thompson *et al.*, *Nucleic Acids Res.*, 22:4673-4680 (1994)) with default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Default parameters for pairwise alignments using the Clustal method are: KTUPLE =1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "mL" means milliliters, "L" means liters, "μL" means microliters, "g" means grams, "mg" means milligrams, "μg" means micrograms, "ng" means nanograms, "nm" means nanometer, "M" means molar, "mM" means millimolar, and "μM" mean micromolar.

1. Enzymatic synthesis and purification of pHCACoA, the substrate for HCHL enzyme assays.

Expression Cloning of pHCA-CoA Ligase and Recombinant Production of pHCA-CoA Ligase

Measuring hydroxycinnamoyl hydratase/lyase (HCHL) activity in plant extracts and of recombinantly produced enzyme requires pHCACoA, a chemical that is not commercially available. pHCACoA was synthesized enzymatically using a recombinantly produced pHCACoA-ligase enzyme from *Arabidopsis thaliana* (*At4CL1*, GenBank® U18675) and purified by preparative chromatography on C18 reverse-phase cartridges. Briefly, a cDNA clone (acs1c.pk003.m10) was identified in DuPont's expressed sequence tag (EST) database that corresponds to a full-length clone of the *At4CL1* transcript. Two primers Primer 1 ACTATTTCATATGGCGCCACAAGAACAAG (SEQ ID NO:1) and Primer 2: GGTTGAAATCAAGCTTCACAATCCCATTTG (SEQ ID NO:2) were used to amplify an open reading frame that is flanked by *Nde*I and *Hind*III restriction sites for cloning into the *E. coli* expression vector pET28A. The resulting construct expresses a variant of the 4CL1 protein that has an N-terminal hexa-histidine tag. The plasmid construct was introduced into BL21DE3 cells (Invitrogen, Carlsbad, CA) and recombinant protein production was induced under standard conditions at 27 °C by adding IPTG (0.2 mM final concentration). pHCACoA ligase activity was extracted and measured spectrophotometrically as described by Gross *et al.* (*Biochemie und Physiologie der Pflanzen*, 168(1-4):41-51 (1975)). Specific pHCACoA ligase activity of cell free extract of *E. coli* cells (36 mg/mL protein) was 28.6 nkat/mg protein. The extract was supplemented with glycerol (7.5 % final concentration), stored at –80 °C, and used for preparative pHCACoA synthesis without further purification.

Preparative Synthesis and Purification of pHCACoA

Preparative synthesis of pHCACoA was carried out at 30 °C in aliquots of 10 mL in the presence of 0.3 mM free CoA (Sigma, USA), 5 mM ATP, 0.5 mM pHCA, 0.2 M Mops (pH 7.5), and 10 mM $MgCl_2$. Enzymatic synthesis was

started by addition of 600 μL cell-free *E. coli* extract (22 mg protein and 630 nkat of At4CL1 enzyme). Formation of pHCACoA was monitored by HPLC analysis pHCA was detected at λ=290 nm and pHCACoA at λ=335 nm. After 15 min, quantitative conversion of pHCA to pHCACoA was achieved. pHCACoA was purified using C18 reverse-phase cartridges (900 mg resin, Burdick and Jackson, USA) hooked up to a Pharmacia FPLC system (Amersham, USA). Fifty milliliters, equaling five combined enzyme reaction mixtures, were loaded onto the cartridge. The cartridge was washed with 30 mL of 0.2 M Mops (pH 7.5) and pHCACoA was eluted with 20 % MeOH. Fractions containing pHCACoA were identified visually. pHCACoA is bright yellow. Fractions were pooled, lyophilized, and resuspended in 5 mL of 10 mM ammonium acetate (pH 4.7). pHCACoA was quantitated spectrophotometrically using the published the molar absorption coefficient of 21 mM^{-1}. The pHCACoA concentration in the resuspended, lyophilized sample was 3.2 mM, thus this method yielded about 15 mg of pHCACoA. pHCACoA was divided into 100 μL aliquots and stored at -80 °C.

2. HCHL Enzyme Assays

The standard HCHL assay was comprised of 100 mM Tris/HCL (pH 8.5), 0.25-0.5 mM pHCACoA, and enzyme sample (2.5-25 μg of total plant protein, 2.5-20 ng of purified HCHL enzyme) in a final volume of 25 μL. Assays were conducted at 30 °C and stopped by adding of an equal volume of 12 % acetic acid in methanol. Formation of *p*-hydroxybenzaldehyde (pHBALD) from pHCACoA in the enzyme assay was measured by HPLC analysis. The reaction mixture was cleared by centrifugation. Reaction products (10 μL) were injected onto a Nova Pak C18 column (3.9x150 mm, 60Å, 4 μm) (Waters, MA, USA). The column was developed at a flow-rate of 1 mL/min under the following conditions: Solvent A (H_2O,/1.5 % HPO_4), Solvent B (50 % MeOH/ H_2O/1.5% HPO_4); 0-5min 0% B, 5-20min 0-100 % B (linear gradient), 20-21min 100-0 % B, and 21-25min 0 % B. pHBALD was detected at 283 nm and quantitated using standard curves established by HPLC separation of known concentration of commercially-available pHBALD (Sigma, USA).

3. Plant growth and transformation

Plant Growth

If not stated otherwise, plants were grown under standard conditions (14 h light, 12 h darkness) in a greenhouse. Plants expressing HCHL genes where grown at 100 μE m^{-2} sec^{-1}, 14 h light (23 °C), 12 h (18 °C) darkness and 70 % relative humidity in growth chambers (Conviron, USA). Sterile plant

cultures were maintained under identical conditions in a plant growth chamber (Percival, USA).

Plant Transformation

Arabidopsis thaliana plants were transformed using Agrobacterium strains (C58, C1 GV3101 MP90) (Koncz, C. and Schell, J., *Mol. Gen. Genet.*, 204:383-396 (1986)) and published protocols of the *in-planta* transformation method (Desfeux *et al.*, *Plant Physiology*, 123(3):895-904 (2000)). Selection for transformants carrying the *NPTII* gene was conducted on sterile growth media in the presence of 50 mg/L kanamycin. Selection for transformants carrying the *BAR* gene was conducted on sterile growth media in the presence of 7.5 mg/L glufosinate or by germinating seed in soil followed by spray-application of an aqueous solution (6 mg/L) of glufosinate herbicide (Sigma, USA) 7 days after germination. Plants destined for plant transformation experiments were grown under permanent light at 23 °C to accelerate flower development.

4. pHBA analysis

pHBA was quantitated in plant tissue by HPLC analysis. For determination of pHBA conjugates, fresh oven-dried or lyophilized tissue was extracted with 50 % MeOH. To quantitate free pHBA plant samples (fresh, dried, lyophilized plant tissue or dried-down methanol extracts of plant tissue) were subjected to acid hydrolysis. Dried or lyophilized tissue was ground to a fine powder using a Cyclotec 1093 tissue mill (Foss Tecator, Sweden) prior to hydrolysis. Tissue (5-25 mg of dried or lyophilized material, 10-100 mg of fresh tissue) was supplemented with 500-750 μL of 1M HCl and incubated at 100 °C for 1-3 h. The hydrolysate was adjusted to alkaline pH by addition of one volume of 1.1 M NaOH. The hydrolysate was cleared by centrifugation and/or filtration and analyzed by HPLC as described above. pHBA or pHBA conjugates were detected at 254 nm and quantitated using standard curves established by HPLC separation of known concentration of commercially-available pHBA or chemically synthesized pHBA conjugates.

EXAMPLES

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Cloning and characterization of an *HCHL* gene from *Pseudomonas putida* (DSM 12585)

Evaluation of HCHL-mediated pHBA production in Arabidopsis focused on the *HCHL* gene from *Pseudomonas putida* (DSM 12585). Muheim and Lerch (*Appl Microbiol Biotechnol*, 51:456-461(1999)) reported that this strain was able to convert ferulic acid to vanillin and several studies have reported the cloning of an *HCHL* gene from closely related *Pseudomonas* strains encoding the HCHL enzyme that is responsible for this activity. The *Pseudomonas* strain described by Muheim and Lerch (*supra*) was ordered from the DSM (Deutsche Sammlung von Microorganismen und Zellkulturen, Braunschweig, Germany). The strain was able to grow on minimal media (Miller, J; Experiments in Molecular Genetics, 1972, Cold Spring Harbour Laboratory Press) containing 1-10 mM pHCA as sole carbon source providing strong support for the presence of an HCHL enzyme in this organism. Genomic DNA was isolated from this strain using standard methods (Maniatis, *supra*) and used as template in a PCR reaction. Two oligonucleotide primers Primer 3: CCATGAGCACATACGAAGGTCGCTGG, (SEQ ID NO:3) and Primer 4: TCAGCGCTTGATGGCTTGCAGGCC (SEQ ID NO:4) were used to generate a PCR fragment of approximately 900 bp that was cloned into *Eco*RV linearized pSKII+ (Stratagene, CA, USA)) that had been modified for cloning of PCR products. Eight independent plasmid clones were recovered and sequenced. BLAST analysis revealed that consensus nucleotide sequence (SEQ ID NO:5) and deduced amino acid sequence (SEQ ID NO:6) of the *HCHL* gene of *Pseudomonas putida* (DSM 12585) shared 88% and 93 % identity to *HCHL* gene and protein of *Pseudomonas fluorescens* AN103 (GenBank® Y13067), respectively.

Expression Cloning, Purification and Determination of Kinetic Properties

Two primers, Primer 5: CATATGAGCACATACGAAGGTCGC (SEQ ID NO:7), and Primer 6 AAGCTTCAGCGCTTGATGGCTTGCAGG (SEQ ID NO:8) and DNA from the plasmid containing the *HCHL* gene of *Pseudomonas putida* (DSM 12585) were used to amplify an open reading frame that is flanked by *Nde*I and *Hind*III restriction sites for cloning into the *E. coli* expression vector pET29A (Novagen, USA). PCR products were cloned into pSKII+. The *HCHL* gene expression cassette was excised and ligated to *Nde*I and *Hind*III-digested pET29a DNA. Amino acid sequence of the HCHL protein expressed from the pET29a HCHL construct is identical to that set forth as SEQ ID NO:6. A second expression construct was generated that expresses a variant of the HCHL

protein that carries a C-terminal hexa-histidine tag. Two primers, Primer 5 (SEQ ID NO:7) and Primer 7: AAGCTTGCGCTTGATGGCTTGCAG (SEQ ID NO:9), and DNA from a plasmid containing the *HCHL* gene of *Pseudomonas putida* (DSM 12585) were used to amplify an open reading frame that is flanked by *Nde*I and *Hind*III restriction sites for cloning into the *E. coli* expression vector pET29A. PCR products were cloned into pSKII+. The HCHL gene expression cassette was excised and ligated to *Nde*I, *Hind*III-digested pET29a DNA. Amino acid sequence of the HCHL protein expressed from the pET29a HCHL 6xHis Tag construct is set forth as SEQ ID NO:10.

Purification and Kinetic Properties of the His-Tagged HCHL Enzyme from *Pseudomonas putida* (DSM 12585)

LB medium (200 mL containing 50 mg/L kanamycin) was inoculated with a single colony of *E. coli* BL21DE3 cells harboring the pET29a HCHL 6xHis Tag expression construct. Cells were grown to an $OD_{\lambda=600\ nm}$ of 0.6 and protein production was induced by addition of 0.2 mM IPTG. Cells were grown at room temperature for 24 h. Cells were harvested by centrifugation (5000xg for 10min) and resuspended in 2.5 mL of 100 mM Tris/HCl (pH 8.5), 20 mM DTT, and 300 mM NaCl. The cell suspension was passed twice through a French press and cleared by centrifugation (30000xg, 20 min, at 4 °C). The cell-free extract was buffer-exchanged using PD10 columns into 20 mM $NaPO_4$ (pH 7.5), 500 mM NaCl, and 10 mM imidazole and loaded on a 5 mL HiTrap chelating chromatography cartridge (Amersham Pharmacia, USA). The column was washed with 20 mL of loading buffer and 20 mL of loading buffer containing 70 mM imidazole. The his-tagged HCHL protein was eluted from the column with a linear gradient from 70-1000 mM imidazole in loading buffer.

HCHL activity in the fractions was determined using a visual assay. Briefly, 0.5 µL of chromatography fractions were added to 25 µL of an HCHL reaction mix (see general methods) that contained feruloylCoA. In the presence of HCHL enzyme activity, the yellow feruloylCoA was rapidly converted to vanillin, which is accompanied by a disappearance of color. Two 1-mL fractions with HCHL activity were pooled and desalted into HCHL extraction buffer. Visual inspection of Coomassie-stained PAGE gels indicated that the HCHL enzyme was greater than 95 % pure. HCHL enzyme concentration was determined spectrophotometrically using an extinction coefficient of 54,600 M^{-1} at 280 nm as determined by the GCG Peptidesort program using the amino acid composition of the his-tagged enzyme variant. The final concentration of the purified recombinant HCHL protein was 2.077 mg/mL, which corresponds to a monomer concentration of 64.139 µM and a concentration of active sites of

32.069 μM. Remaining fractions with HCHL activity were pooled and quantitated in a similar fashion. HCHL protein (17 mg) was purified from 250 mg of total *E. coli* protein indicating that the recombinant protein represented at least 7 % of the total protein. Kinetic properties of the HCHL enzyme were determined. Standard HCHL assays were conducted using pHCACoA and feruloylCoA concentrations ranging from 343 to 2.7 μM and 293 to 2.3 μM, respectively. Assays were incubated for 5.5 min and pHBALD and vanillin were quantitated by HPLC.

The Michaelis-Menten and Wolf-Augustinsson-Hofstee plots (Figure 2) of the data indicate that the Km and Vmax values of the his-tagged HCHL enzyme from *Pseudomonas putida* for pHCACoA and feruloylCoA were 2.53 μM, 53.8 nkat/mg, and 2.39 μM, 37.3 nkat/mg, respectively. The Vmax of the enzyme with pHCACoA translates into a catalytic center activity of 3.4/sec (per enzyme dimer), which was calculated using a molecular weight of 32,348.8 Da per monomer. This is in very close agreement with the published Vmax and Km values of the HCHL enzyme from *Pseudomonas fluorescens* AN103 (Mitra *et al.*, *Arch. Biochem. Biophys.*, 365(1):6-10 (1999)). The values were reported to be 5.3 μM, 73 nkat/mg for pHCACoA and 2.4 μM, 36.5 nkat/mg for feruloylCoA.

Purification and Kinetic Properties of the Native HCHL Enzyme from *Pseudomonas putida* (DSM 12585)

LB medium (500 mL containing 50 mg/L) kanamycin was inoculated with a single colony of *E. coli* BL21DE3 cells harboring the pET HCHL expression construct. Cells were grown to an $OD_{\lambda=600\ nm}$ of 0.6 and protein production was induced by the adding of 0.2 mM IPTG. Cells were grown at room temperature for 24 h. Cells were harvested by centrifugation (5000xg for 10 min) and resuspended in 15 mL of 100 mM Tris/HCl (pH 8.5), 20 mM DTT, and 300 mM NaCl. The cell suspension was passed twice through a French press and cleared by centrifugation (30000xg, 20 min, at 4 °C) resulting in 15 mL of cell-free extract (38.3 mg/mL protein). Two 2.5-mL aliquots of the cell-free extract were buffer-exchanged using PD10 columns into 50 mM Tris/HCl (pH 7.6), 10 mM Na_2SO_3, and 1 mM EDTA. Buffer-exchanged extract (7 mL) was loaded onto a Q-sepharose column (15 mL gel bed volume). The column was developed at a flow rate of 4 mL/min at 4 °C as follows: Solvent A (50 mM Tris/HCl (pH 7.6), 10 mM Na_2SO_4, and 1 mM EDTA), Solvent B (1 M NaCl, 50 mM Tris/HCl (pH 7.6), 10 mM Na_2SO_3, and 1 mM EDTA); 0-20 min 0% B, 20-80 min (linear gradient) 0-100% B, 80-100 min 100 % B, and 101-120 min 0% B. Fractions (4 mL) were collected and HCHL activity was monitored visually as described previously. A fraction with HCHL activity was purified further by

chromatography on hydroxyapatide (Biorad Econo-Pac cartridge CHT-II, 1 mL gel bed volume (Biorad, CA, USA)). Approximately 2.5 mL was buffer-exchanged into (10 mM $NaPO_4$, (pH 6.8) and 10 μM $CaCl_2$). The column was developed at a flow rate of 2 mL/min at 4 °C as follows: Solvent A (10 mM $NaPO_4$, (pH 6.8) and 10 μM $CaCl_2$), Solvent B (350 mM $NaPO_4$, (pH 6.8) and 10 μM $CaCl_2$); 0-10 min 0 % B, 10-30 min (linear gradient) 0-100 % B, 30-50 min 100 % B, and 51-70 min 0 % B. Fractions (1 mL) were collected, assayed for HCHL activity as described above, and analyzed by PAGE. Visual inspection of Coomassie-stained gels indicated that in some chromatography fractions the HCHL enzyme was greater than 90 % pure. HCHL enzyme concentration was determined spectrophotometrically using an extinction coefficient of 54,600 M^{-1} at 280 nm as determined by the GCG Peptidesort program using the amino acid composition of the native enzyme. The final concentration of the purified recombinant HCHL protein in this fraction was 0.311 mg/mL, which corresponds to a monomer concentration of 10.073 μM and a concentration of active sites of 5.04 μM.

The Michaelis-Menten and Wolf-Augustinsson-Hofstee plots of the data indicate Km and Vmax values of the native HCHL enzyme from *Pseudomonas putida* for pHCACoA were 2.4 μM and 43 nkat/mg' respectively. The Vmax of the enzyme with pHCACoA translates into a catalytic center activity of 2.65/sec (per enzyme dimer) that was calculated using a molecular weight of 30,865.1 Da per monomer. This is in close agreement with the published Vmax and Km values of the HCHL enzyme of *Pseudomonas fluorescens* AN103 (Mitra *et al.*, *supra*). The kinetic properties of *the Pseudomonas putida* HCHL enzyme for conversion of pHCACoA to pHBALD did not deviate significantly from values published for the HCHL enzyme of *Pseudomonas fluorescens* AN103.

EXAMPLE 2

Plant Expression of the *HCHL* Gene of *Pseudomonas putida* (DSM 12585) Under the Control of Constitutive Promoters.

Construction of binary vectors

For constitutive expression of the *Pseudomonas putida* (DSM 12585) HCHL enzyme in plants, two binary vectors were generated. In one construct, the *HCHL* gene was under the control of the promoter of the *ACTIN2* gene from Arabidopsis. It had been shown previously that this promoter confers a constitutive pattern of reporter gene expression in plants (An *et al.*, *Plant Journal*, 10(1):107-121 (1996)). In the other construct, the HCHL coding sequence was fused to the *CaMV35S* promoter.

Genomic DNA from *Arabidopsis thaliana* plants and two PCR primers used to amplify a 1220 bp fragment of the *ACTIN2* gene that comprised promoter region and 5'UTR of the gene: Primer 8 CAACTATTTTTATGTATGCAAGAGTCAGC (SEQ ID NO:11) and Primer 9 CCATGGTTTATGAGCTGCAAACACAC (SEQ ID NO:12). The sequence of the *ACTIN2* promoter ("*ACT2*") fragment is set forth as SEQ ID NO:13. Primer 9 introduced an *Nco*I site (CCATGG) at the start codon and permitted generation of translational fusions (at the start codon) of the *ACT2* promoter to any gene of interest that has been modified to carry an *Nco*I or *Pag*I site at the start codon. The *ACT2* promoter fragment was cloned into an *Eco*RV linearized pSKII+ vector modified for cloning of PCR products. Four plasmid clones were recovered in which the 3' end of the promoter was proximal to the *T7* promoter in the pSKII+ vector. The *ACT2* promoter was released from the vector using the restriction enzymes *Hind*III and *Nco*I.

The *P. putida* (DSM 12585) *HCHL* coding sequence was amplified from the plasmid template (see above) using two primers: Primer 10: TCATGAGCACATACGAAGGTCGC (SEQ ID NO:14) and Primer 4 (SEQ ID NO:4). Primer 10 introduced a *Pag*I (TCATGA) at the start codon and facilitated the fusion of the *HCHL* coding sequence to the *ACT2* promoter. The PCR products were cloned into pSKII+ and two clones were recovered in which the start codon of the *HCHL* coding sequence is proximal to the to the *T7* promoter in the pSKII+ vector. Plasmid DNA of these clones was linearized by partial digestion with *Pag*1 and the *HCHL* coding sequence was released from the vector by complete digestion with *Sst*I. The *HCHL* coding sequence and *ACT2* promoter were assembled in a three-way ligation to *Hind*III and *Sst*I digested pSKII+ vector DNA. The *ACT2-HCHL* expression cassette was excised and ligated to *Hind*III/*Sst*I digested DNA of the binary vector pGPTVBar (Becker *et al.*, *Plant Molecular Biology*, 20(6):1195-7 (1992)). Recombinant Plasmid DNA was isolated from *E. coli* and introduced into *Agrobacterium tumefaciens* for transformation of wild type Arabidopsis plants.

A *CaMV35S* promoter with a duplicated enhancer element was excised from the plasmid pJIT60 (Transformation of *Brassica oleracea* with paraquat detoxification gene(s) mediated by *Agrobacterium tumefaciens*. Latifah, A.; Salleh, M.A.; Basiran, M. N.; Karim, A.G. Abdul. Faculty Life Sciences, University Kabangsaan Malaysia, Malay. Editor(s): Shamann, Nor Aripin. Applications of Plant In Vitro Technology, Proceedings of the International Symposium, Serdang, Malay., Nov. 16-18, 1993 (1993), 145-50) using restriction digestion with *Kpn*I and *Hind*III and cloned into pSKII+. The modified

pSKII+ vector was digested with *Eco*RV and T-tailed for cloning of PCR products using Taq polymerase. The *HCHL* coding sequence of *P. putida* was amplified from a pSKII+ plasmid template using Primer 10 (SEQ ID NO:14) and Primer 4 (SEQ ID NO:4) and inserted downstream of the *CaMV35S* ("*35S*") promoter in the modified pSKII+ vector. Four plasmid clones were recovered in which the start codon of the *HCHL* coding sequence is proximal to the 3' end of the *CaMV35S* promoter. Insert DNA was excised from these plasmids by digestion with *Kpn*I/*Sst*I and ligated to pGEM7zf+ (Promega, USA) digested with the same enzymes. This cloning step introduced an *Xba*I site at the 5' end of the *35S HCHL* expression cassette. The pGEM7zf+ construct was linearized with *Xba*I by partial digestion. The *HCHL* expression cassette was released from the vector by complete digestion with *Sst*I. The *35S-HCHL* expression cassette was ligated to *Xba*I/*Sst*I digested DNA of the binary vector pGPTVBar (Becker *et al.*, *supra*). Recombinant Plasmid DNA was used for transformation of wild type Arabidopsis plants as described in general methods.

Analysis of pHBA levels in leaves of primary transformants

Act2 HCHL: 105 primary transformants were identified based on their ability to survive application of the glufosinate herbicide. These transformants were grown in soil for 28 days. pHBA content of leaf tissue was determined by HPLC analysis as described in the general methods. pHBA content in leaf tissue of the primary transformants ranged from 0.59 to 5.47 mg/g DW. One line (119) was self-crossed and T2 seed were harvested. Segregation analysis of the selectable marker was conducted at the T2 level and seed batches homozygous for the T-DNA insertion were identified in the T3 generation. Homozygous seed material of this line was used for subsequent experimentation.

CaMV35S HCHL: 16 primary transformants were identified based on their ability to survive application of the glufosinate herbicide. These transformants were grown in soil for 28 days. pHBA content of leaf tissue was determined by HPLC analysis as described in the general methods. pHBA content in leaf tissue of the primary transformants ranged from 0.95 to 7.69 mg/g DW. One line (11) was self-crossed and T2 seed were harvested. Segregation analysis of the selectable marker was conducted at the T2 level and seed batches homozygous for the T-DNA insertion were identified in the T3 generation. Homozygous seed material of this line was used for subsequent experimentation.

Substrate Limitation in Leaf tissue

To gain insights into the limitations of HCHL-mediated pHBA production in leaf tissue, wild type Arabidopsis plants and homozygous plants of lines 11

and 119 were grown in soil. Leaf material was harvested six weeks after germination. Concentrations pHBA and sinapic acid were determined by HPLC analysis.

In leaf tissues of Arabidopsis the substrate of HCHL, pHCACoA, is used as an intermediate for synthesis of aromatic secondary metabolites such as flavonoids and UV-fluorescent sinapic acid esters. The accumulation of the latter in the cells of the upper leaf epidermis endows the Arabidopsis leaves with a characteristic green-blue fluorescence under long wave UV light. Leaves of wild type and transgenic lines expressing the *HCHL* gene were illuminated with long wave UV light (λ= 366nm). Applicants observed a red fluorescence under long wave UV light of leaves of transgenic lines 11 and 119. This indicates the depletion of sinapate esters as result of HCHL expression. This conclusion was further confirmed by HPLC analysis (Table 3) demonstrating that formation of pHBA from pHCACoA by HCHL is accompanied by a significant depletion of sinapic acid. This result indicates that, in leaf tissue, formation of pHCACoA limits the rate of pHBA synthesis by HCHL. In other words, HCHL is operating in substrate-limited mode in leaf tissue. It is interesting to note that in the best HCHL expressing line (11) the observed level of pHBA accumulation is achieved through a five-fold increase of flux through the phenylpropanoid pathway when compared to wild type plants. This corroborates findings by Mayer *et al.* (*supra*), indicating that an increase of steady-state transcript levels of genes such as *PAL*, *C4H,* and *4CL* accompany expression of an *HCHL* gene in transgenic tobacco.

Table 3

Construct	Sinapic acid (µmol/g FW)	pHBA (µmol/g FW)
WT	1.65	0
Act2 HCHL (119)	0.71	4.32
35S HCHL (11)	0.09	8.71

Leaves versus Stems

The next objective was to investigate the efficaciousness of the HCHL route of pHBA production in stalk tissue. In this tissue the HCHL substrate, pHCACoA, is a central intermediate of a high flux pathway that provides precursors for lignin biosynthesis shown in Figure 1. The high flux nature of this pathway is illustrated by the fact that even in herbaceous plants, such as

Arabidopsis, lignin constitutes approximately 20 % of the dry matter of the stalk tissue.

Homozygous transgenic lines 119 and 11 were grown in soil for 8 weeks. Leaf and stalk tissue was harvested, lyophilized, and ground to a powder that was subjected to analysis of pHBA content by acid hydrolysis and HPLC. In transgenic lines constitutively expressing *HCHL*, pHBA accumulation in the stalk tissue was dramatically higher in stalk tissue in comparison to leaf tissue. pHBA levels of 18.3 mg/g DW and 6.9 mg/g DW were detected in whole stalk tissue from lines 11 and 119, respectively. This is significantly higher than 13 mg/g DW and 3.8 mg/g DW detected in leaf tissue of the same lines.

In order to confirm that the high impact of HCHL on pHBA production in stalk tissue reflected substrate availability and not enzyme activity in this tissue, leaf and stalk tissue of line 11 (*35S HCHL*) was assayed for HCHL enzyme activity and pHBA content was determined (Table 4). For this experiment the basal stem segment was used. Table 4 shows that although HCHL enzyme activity differs only by 60 % when leaf and stalk tissue are compared, pHBA content is 6-fold higher in stalk tissue.

Table 4

Line	Tissue	HCHL activity (pkat/mg)	pHBA (mg/g DW)
35S HCHL (11)	**leaves**	**100**	**4.6**
	stems	**160**	**30.5**

Correlation of HCHL enzyme activity and pHBA accumulation in stalk tissue

As a prelude to work on further improvements of HCHL-mediated pHBA production in stalk tissue, Applicants investigated whether there is any indication of substrate limitation of HCHL in stalk tissue of the transgenic lines generated so far. T2 seed material of different transgenic Arabidopsis lines expressing the *35S HCHL* transgene was germinated on phosphinotrine-containing growth media and herbicide-resistant plants were grown in soil for eight weeks. Stalk tissue was harvested and subjected to pHBA analysis and HCHL assays. HCHL transformants were selected for this experiment that covered a wide range of pHBA accumulation in leaf tissue of primary transformants. Figure 3 shows a linear correlation (R^2=0.8261) between specific HCHL activity and pHBA content over a wide range of specific HCHL activity, indicating that in the lines with the highest specific HCHL activity in stalk tissue there is no indication of substrate limitation.

EXAMPLE 3

Stalk-specific Expression of HCHL in Plants

In this Example, the utility of different stalk-specific promoters was determined. A pattern of *HCHL* expression that targeted the specialized cell types having a high rate of pHCACoA synthesis would produce a high level of pHBA in the stalk. Lignin biosynthesis is a cell autonomous process. RNA blot experiments, expression of reporter gene constructs, and immunolocalization studies of enzymes of the phenylpropanoid pathway suggest that the bulk of monolignols is produced in the cells that undergo lignification. There is only a limited transfer of monolignols from neighboring xylem or ray parenchyma cells to tracheids or vessel elements (presumably at later stages of cell differentiation) to sustain lignification after the water-conducting cells have undergone autolysis. The promoters of genes closely related to the synthesis of (or consumption of) pHCACoA in lignin biosynthesis were selected in order to target *HCHL* expression to plant stalk tissue. The goal was to identify those promoters that would lead to pHBA accumulation, in excess of the the levels observed with a constitutive promoter, such as *CaMV35S*, by targeting the cells with the highest concentration of the pHCACoA substrate. Successful targeting of HCHL to these cell types was expected to avoid the detrimental effects associated with depleting of pHCACoA in tissues other than stalk tissue.

Construction of Plasmids for Expression of *HCHL* in Plants Under Control of *C4H*, *4CL1*, *C3'H*, and *IRX3* Promoters.

Cinnamate-4-hydroxylase (C4H) catalyzes the 4-hydroxylation of the aromatic ring of cinnamic acid. C4H (CYP73A5; GenBank® Accession No. U71080) is a cytochrome P450-dependent monooxygenase encoded by a single gene in most plants. Genomic DNA was isolated from Arabidopsis plants and the primers Primer 11: GAGAGCATCCATATGAGCACATACGAAGGTCGC (SEQ ID NO:15) and Primer 12: CGCAGCGTCAAGCTTCAGCGTTTATACGCTTGC (SEQ ID NO:16) were used to amplify 2721 nucleotides of the *C4H* promoter (SEQ ID NO:17). PCR products were cloned into the pCR2.1 vector (Invitrogen, USA). Primer 12 introduces a *Nco*I site (CCATGG) at the initiator methionine codon of the *C4HL gene* and facilitates the generation of translational fusions of genes that contain *Pag*I (TCATGA) or *Nco*I sites at the start codon. A pSKII+ plasmid containing a PCR-generated variant of the *HCHL* gene containing a *Pag*I site at the start codon was partially digested with *Pag*I and a *Pag*I/*Sst*I fragment was released from the vector by complete digestion with *Sst*I. The *C4H* promoter was released from the pCR2.1 vector by digestion with *Xba*I/*Pag*I. The *C4H* promoter

and *HCHL* gene were assembled in the *Xba*I-*Sst*I cut pGPTVBar vector (Becker *et al.*, *supra*) in a three-way ligation. Plasmid DNA was used for agrobacterium-mediated transformation of Arabidopsis plants.

4-Coumarate-coenzymeA ligase (4CL) enzymes are operationally soluble, monomeric enzymes of 60 kDa molecular weight belonging to the class of adenylate forming CoA ligases. There is clearly redundancy at the level of 4CL enzyme activity both in gymnosperms and angiosperms. Angiosperm 4CL proteins belong to two groups of evolutionarily divergent sequences. For example, in Arabidopsis there are three distinct 4CL proteins that share only 60 % sequence identity. The *4CL1* (GenBank® Accession No. U18675) gene is constitutively and abundantly expressed in plant stem tissue, indicating that it carries out an important role in lignin biosynthesis. In contrast, the expression pattern of the *4CL2* and *4CL3* genes are expressed in response to environmental cues and is also observed in tissues other than the stalk (Ehlting *et al.*, *PLANT JOURNAL*, 19(1):9-20 (1999)).

Genomic DNA was isolated from Arabidopsis plants and the primers Primer 13: CCTAGAAGTGTTGCAGCTGAAGGTACTAAC (SEQ ID NO:18) and Primer 14: GTTCTTGTGGCGCCATGGTAAATAGTAAAT (SEQ ID NO:19) were used to amplify 2739 nucleotides of the *4CL1* promoter (SEQ ID NO:20). PCR products were cloned into the pCR2.1 vector. Primer 14 introduced an *Nco*I site (CCATGG) at the initiator methionine codon of the *4CL1* gene and facilitated the generation of translational fusions of genes that contain *Pag*I (TCATGA) or *Nco*I sites at the start codon. A pSKII+ plasmid containing a PCR-generated variant of the *HCHL* gene containing a *Pag*I site at the start codon was partially digested with *Pag*I and a *Pag*I/*Sst*I fragment was released from the vector by complete digestion with *Sst*I. The *4CL1* promoter was released from the pCR2.1 vector by digestion with *Xba*I/*Pag*I. The *4CL1* promoter and *HCHL* gene were assembled in the *Xba*I-*Sst*I cut pGPTVBar vector in a three-way ligation. Plasmid DNA was used for agrobacterium-mediated transformation of Arabidopsis plants.

The p-coumarate-3-hydroxylase gene (*C3'H*) encodes a 3-hydroxylase enzyme (CYP98A3, GenBank® Accession No. AC011765) that generates the 3,4-hydroxylated caffeoyl intermediate in lignin biosynthesis. Characterization of the kinetic properties and substrate specificity of this enzyme revealed that shikimate and quinate esters of the 4-hydroxylated coumaryl moiety constitute the preferred substrate of the 3-hydroxylase (Schoch *et al.*, *J Biol. Chem.*, 276(37):36566-36574 (2001)).

Genomic DNA was isolated from Arabidopsis plants and the primers Primer 15: CGATTTTGATCGTTGACTAGCTATACAATCCC (SEQ ID NO:21)

and Primer 16: GCTATTAGAAACCACGCCATGGAGTTTTGCTTC (SEQ ID NO:22) were used to amplify 2705 nucleotides of the *C3'H* promoter (SEQ ID NO:23). PCR products were cloned into the pCR2.1 vector. Primer 16 introduces a *Nco*I site (CCATGG) at the initiator methionine codon of the *C3'H* gene and thus facilitates the generation of translational fusions of genes that contain *Pag*I (TCATGA) or *Nco*I sites at the start codon. A pSKII+ plasmid containing a PCR-generated variant of the *HCHL* gene containing a *Pag*I site at the start codon was partially digested with *Pag*I and a *Pag*I/*Sst*I fragment was released from the vector by complete digestion with *Sst*I. The *C3'H* promoter was released from the pCR2.1 vector by partial digestion with *Xba*I and complete digestion with *Pag*I. The *C3'H* promoter and *HCHL* coding sequence were assembled in the *Xba*I-*Sst*I cut pGPTVBar vector in a three-way ligation. Plasmid DNA was used for agrobacterium-mediated transformation of Arabidopsis plants.

The *IRX3* (irregular xylem 3) gene of Arabidopsis encodes one of the catalytic subunits comprising the cellulose synthesis catalytic complex (*AtCESA7*, GenBank® Accession No. AF091713) that is essential for cellulose synthesis in stalk tissue (Turner *et al.*, *Plant Cell*, 9(5):689-701 (1997); Taylor *et al.*, *supra* (1999); and Taylor *et al.*, *supra* (2003)). The corresponding wild type version of this gene is denoted at *AtCesA7*. The role of this gene in forming of the plant stalk was revealed by genetic analysis. A mutation in this gene almost completely abolishes cellulose deposition in secondary cell walls in the stalk, but does not affect cellulose deposition in primary cell walls and other tissues of the plant. The promoter of this gene has been employed for down-regulation of enzymes involved in lignin biosynthesis (Jones *et al.*, *Plant Journal*, 26(2):205-216 (2001)). Although the *AtCesA7* (*IRX3*) gene product does not have a role in lignin biosynthesis, it controls a process that is closely associated with lignin deposition in the secondary cell walls of the stalk. The *AtCesA7* (*IRX3*) promoter was evaluated for its utility in targeting *HCHL* expression to the plant stalk.

Genomic DNA was isolated from Arabidopsis plants and the primers Primer 17: CAGTTTATCTGGGTAAGTTCTTGATTTTAAGC (SEQ ID NO:24) and Primer 18: GACCGGCGCTAGCTTTCATGAGGACGGCCGGAG (SEQ ID NO:25) were used to amplify 2780 nucleotides of the *AtCesA7* (*IRX3*) promoter. PCR products were cloned into a pCR2.1 vector. Primer 18 introduced a *Pag*I site (TCATGA) at the initiator methionine codon of the *AtCesA7* (*IRX3*) gene and facilitated the generation of translational fusions of genes that contain *Pag*I (TCATGA) or *Nco*I sites at the start codon. A pSKII+ plasmid containing a PCR-

generated variant of the *HCHL* gene containing a *Pag*I site at the start codon was partially digested with *Pag*I and a *Pag*I/*Sst*I fragment was released from the vector by complete digestion with *Sst*I. A 2134 bp fragment (SEQ ID NO:26) of the *AtCesA7* (*IRX3*) promoter was released from the pCR2.1 vector by digestion with *Xba*I and *Pag*I. The *AtCesA7* (*IRX3*) promoter and *HCHL* gene were assembled in the *Xba*I-*Sst*I cut pGPTVBar vector in a three-way ligation. Plasmid DNA was used for agrobacterium-mediated transformation of Arabidopsis plants.

Sequences of fusion products between the *HCHL* gene from *Pseudomonas putida* (DSM 12585) and the promoters from the *C4H*, *4CL1*, *C3'H*, and *AtCesA7* (*IRX3*) genes of *Arabidopsis thaliana* are set forth as SEQ ID NOs:27, 28, 29, and 30, respectively.

Analysis of pHBA in Stalk Tissue of Primary Transformants

Primary transformants were grown in soil for eight weeks. A stem segment of 2 cm was harvested at the base of the stem from each transformant and subjected to analysis of pHBA content by acid hydrolysis and HPLC. Seed material was harvested from the ten best transformants and the remaining stalk material was harvested, dried, ground, and subjected to analysis of pHBA content. Table 5 shows that the *C4H*, *AtCesA7* (*IRX3*), and *4CL1* promoters were able to target HCHL-mediated pHBA production to levels that was comparable to the *CaMV35* promoter. *AtCesA7* (*IRX3*) and *4CL1* lines contained 60 % of the pHBA levels found in the best *35S HCHL* line. pHBA content of whole stalk tissue in the best *C4H HCHL* line was 106 % in comparison to the levels generated by the *35S* line.

Table 5

Construct	n	Basal stalk pHBA average (mg/g FW)	Basal stalk pHBA highest (mg/g FW)	Whole stalk pHBA highest (mg/g DW)
35S HCHL	43	0.82	5.86	22.08 (line 276)
C4H HCHL	78	0.89	5.54	23.42 (line 35)
4CL1 HCHL	71	0.55	3.87	12.93 (line 183)
C3'H HCHL	64	0.41	1.65	9.52 (line 227)
AtCesA7 (IRX3) HCHL	46	1.21	3.91	12.96 (line 366)

Analysis of Whole Stalk pHBA and HCHL Enzyme Activity in Pooled Leaf and Stalk Tissue of T2 lines

The primary transformants were self crossed and T2 seed material was germinated on selective media containing glufosinate, transferred to soil, and grown for eight weeks. Leaf and stalk tissue was harvested and subjected to pHBA analysis and assayed for HCHL activity (Table 6). All promoters provided improved of stalk specificity at the level of HCHL enzyme activity. However, since HCHL runs substrate-limited in leaf tissue, the HCHL activity measured in leaf tissue of the C4H, 4CL1, and C3H lines was still sufficient to convert all available pHCACoA to pHBA. The improved stalk specificity of HCHL expression did not translate into improved stalk specificity of pHBA deposition in these lines. In other words, the three promoters from genes involved in lignin biosynthesis (*C4H*, *C3'H*, and *4CL1*) permitted significant HCHL expression in leaf tissue.

Leaf tissue from transgenic lines expressing the *HCHL* gene under the control of the *AtCesA7* (*IRX3*) promoter, on the other hand, exhibited no detectable HCHL activity. pHBA accumulation was reduced more than ten-fold when compared to the *35S HCHL* line. The data indicated that only certain cellulose synthase promoters provide the ideal molecular tools to target HCHL to the plant stalk at levels that can sustain pHBA production comparable to levels achieved with constitutive promoters. The *AtCesA7* (*IRX3*) *HCHL* lines were phenotypically indistinguishable from wild type plants, indicating that restricting of HCHL expression to the plant stalk was compatible with normal plant growth and development.

Table 6

Tissue	Construct	Line	HCHL activity (pkat/mg protein)	PHBA (mg/g DW)	Ratio stem/leaf pHBA	HCHL efficacy (pHBA/pkat/mg)
stem	*35S HCHL*	276	160.2	19.8	3.6	0.12
leaf			65.8	5.5		
stem	*AtCesA7 (IRX3) HCHL*	366	24.6	23.2	43	0.94
leaf			0.0	0.5		
stem	*AtCesA7 (IRX3) HCHL*	365	25.4	13.3	29	0.52
leaf			0.0	0.5		
stem	*C4H HCHL*	35	48.4	15.5	3.8	0.32
leaf			5.3	4.1		
stem	*C4H HCHL*	72	25.8	9.2	2.6	0.36
leaf			1.5	3.5		
stem	*C3'H HCHL*	227	14.7	9.1	4.2	0.61
leaf			1.0	2.1		
stem	*4CL1 HCHL*	140	29.0	19.6	5.5	0.67
leaf			3.1	3.5		

EXAMPLE 4

Isolation of Maize (*Zea mays*) *CesA* cDNA Clones and Amino Acid Sequence Comparisons to Arabidopsis CesA Proteins

Applicants have demonstrated how promoters of certain cellulose synthase genes controlling cellulose deposition in the secondary cell walls of the plant vascular system allow precise targeting of HCHL expression and pHBA production to the plant stalk. Certain grasses (monocotyledonous plants), such as sugar cane, would provide an ideal platform for producing of pHBA in stalk tissue. Not only does the stalk of sugar cane plants provide plentiful biomass, but it also possesses established infrastructure for harvesting and isolating of small water-soluble molecules. We hypothesized genes from monocotyledonous plants that are orthologs *(i.e.,* those that carry out the function of the *AtCesA8 (IRX1), AtCesA7 (IRX3),* and *AtCesA4 (IRX5)* genes of Arabidopsis) would provide promoter sequences suitable for precise targeting of

HCHL expression to the stalk based on the expression pattern reported for these genes in Arabidopsis (Taylor *et al.*, *supra* (2003)).

Holland *et al.* isolated and characterized nine members (*ZmCesA1-ZmCesA9*) of the cellulose synthase gene family of corn (*Zea mays*) (*Plant Physiol.*, 123:1313-1324 (2000) Table 7.). Using methodology described by Holland *et al.* (*supra*), Applicants have isolated three new members of the maize *CesA* gene family (*ZmCesA10*, *ZmCesA11*, and *ZmCesA12*) from the elongation and transition zones of an elongating maize internode. Coding sequences for *Zea mays ZmCesA10, ZmCesA11,* and *ZmCesA12* genes and the corresponding deduced amino acid sequences are provided as SEQ ID NOs:31-36 (Table 7). The DNA upstream of the respective start codon for *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* was sequenced. The respective promoter sequences were identified and are provided as SEQ ID Nos:81, 82, and 83.

Maize and Arabidopsis *CesA* genes were aligned using the CLUSTAL W program (Thompson *et al.*, Nucleic Acids *Res.*, 22:4673-4680(1994)). Protein sequences for the *Arabidopsis* CesA proteins were deduced from the publically available nucleotide sequences in GenBank® (Table 7). Maize sequences for the genes *ZmCesA1* through *ZmCesA12* are available in GenBank® (Table 7; Holland *et al., supra*).

Parsimony and neighbor-joining analyses were performed using the PAUP program (Swofford, DL, PAUP*: Phylogenetic analysis using parsimony (and other methods), Volume Version 4 (Sinauer Associates, Sunderland, MA)). To assess the degree of support for each branch on the tree, bootstrap analysis with 500 replicates was performed (Felsenstein, J., *Evolution*, 39:783-791 (1985)). A maximum-likelihood tree was also reconstructed using proML algorithm implemented in the PHYLIP package by J. Felsenstein (Phylogeny Inference Package, version 3.6a2.1; available from the University of Washington, Seattle, WA). Both neighbor-joining and maximum-likelihood trees showed very similar tree topologies (maximally parsimonious tree with minor terminal branch differences).

The result of this analysis is an unrooted cladogram (Figure 4) comprising the maize and *Arabidopsis* CesA proteins. The deduced amino acid sequences of the maize *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* genes cluster with the corresponding deduced proteins from *Arabidopsis* (*AtCESA4* (*IRX5*), *AtCESA8* (*IRX1*), and *AtCESA7* (*IRX3*), respectively) known to be involved in secondary wall formation. This suggests that the different subclasses of the *CesA* genes diverged early in evolution, at least before monocots and dicots separated (Holland *et al.*, *supra*). Each of the *IRX* genes is expressed in the same cell

type in the vascular tissue in *Arabidopsis* (Taylor *et al.*, *supra* (2003)). Phylogenetic clustering of the maize CesA proteins with the IRX proteins from *Arabidopsis* and the observation that the highest expression was measured in the transition zone of the internode suggest that these genes are involved in secondary wall formation.

Table 7. Genes and Corresponding GenBank® Accession Numbers

Gene Name[1]	GenBank® Accession Number
AtCesA1	AF027172
AtCesA2	AF027173
AtCesA3	AF027174
AtCesA4	AB006703
AtCesA5	AB016893
AtCesA6	AF062485
AtCesA7	AF088917
AtCesA8	AL035526
AtCesA9	AC007019
AtCesA13	AC006300
ZmCesA1	AF200525
ZmCesA2	AF200526
ZmCesA3	AF200527
ZmCesA4	AF200528
ZmCesA5	AF200529
ZmCesA6	AF200530
ZmCesA7	AF200531
ZmCesA8	AF200532
ZmCesA9	AF200533
ZmCesA10 (SEQ ID NOs:31 and 32)	AY372244
ZmCesA11 (SEQ ID NOs:33 and 34)	AY372245
ZmCesA12 (SEQ ID NOs:35 and 36)	AY372246

[1]Source organism represented by first 2 letters of gene name. At = *Arabidopsis thaliana*, Zm = *Zea mays*.

EXAMPLE 5

Expression Analysis of *Zea mays ZmCesA10, ZmCesA11,* and *ZmCesA12* Genes Using MPSS

Expression profiling of the *CesA* gene family: The expression pattern of the maize *CesA* genes in different tissues was studied using the MPSS technology Brenner *et al.*, *Proc. Natl. Acad. Sci. USA*, 97(4):1665-1670 (2000); (Brenner *et al.*, *Nat. Biotech.*, 18:630-634 (2000); Hoth *et al., J. Cell. Sci.,* 115:4891-4900 (2002); Meyers *et al., Plant J.*, 32:77-92 (2002); US 6,265,163; and US 6,511,802). This technology involves attaching each expressed cDNA to the surface of a unique bead. As a result, a highly expressed mRNA is represented on a proportionately large number of beads. Signature sequences of 16-20 nucleotides are then obtained by iteratively restricting the cDNA on a bead with the type IIs endonuclease, adaptor ligation, and hybridizing with an encoded probe. Sequencing of more than a million signatures from each tissue library allows 'electronic Northern' analysis to be carried out. The abundance of a particular mRNA is judged by the ratio of its specific signatures to the total mRNA molecules sequenced and is represented in parts per million (ppm).

Data averaged across multiple libraries for similar tissues (*e.g.*, leaf, stalk, root) are presented in Figure 5. The data are averaged over 76 different libraries. The number of libraries for each tissue was: root, 12; leaf, 13; stalk, 6; ear, 10; silk, 7; kernel, 2; embryo, 10; endosperm, 13; and pericarp, 3. The average for the total number of tags across the 76 libraries was 1,370,525 with a range of 1,223 (721 for a stalk library to 2,154,139 for a root library). The average for the adjusted number of unique tags was 45,293 with a range of 15,226 in an endosperm library to 87,030 for a root library. Similar data from a smaller set of libraries were presented in a previous report (Dhugga, K., *Curr. Opin. Plant Biol.*, 4:488-493 (2001)).

Two general conclusions can be drawn from the data: 1) *CesA* genes 1-8 (with the exception of *CesA*2) are expressed at different levels in a majority of the tissues and 2) *CesA*10-12 are selectively expressed in those tissues that are rich in secondary wall. For *CesA1-8*, the data are in overall agreement with the previously reported data with the exception of *CesA*2, which, after reanalysis is found to be expressed only in the root and the kernel tissues and at a very low level in the silk tissue (Dhugga, K., *supra*). *CesA*5 and *CesA*6 are the highest expressed *CesA* genes in the endosperm and leaf tissues, respectively. *CesA*10, *CesA*11, and *CesA*12 are most highly expressed in the stalk tissue. The expression of none of the *CesA* genes is detected in the mature pollen grain.

Theoretically, the whole expressed genome is analyzed by the MPSS technology each time a library is screened for unique tags. Quantitative measures of the expression levels of different gene tags in the MPSS, as opposed to the ratios across paired tissues or treatments in the microarray-based platforms, combined with the depth of signature sequencing (>1 million) for each of the libraries make it possible to compare gene expression patterns across multiple, independent experiments. A correlation coefficient matrix showing the relationship for the expression pattern among the maize *CesA* genes is shown in Table 8.

Table 8: Correlation coefficient matrix for the expression pattern of maize CesA genes as compiled by the MPSS data. The same data set as used in Figure 5 were used to calculate the correlation coefficients.

Gene Name	Ces A1	Ces A2	Ces A3	Ces A4	Ces A5	Ces A6	Ces A7	Ces A8	Ces A10	Ces A11	Ces A12
CesA1	1.00										
CesA2	0.29	1.00									
CesA3	0.05	-0.08	1.00								
CesA4	0.37	0.17	-0.12	1.00							
CesA5	-0.20	-0.15	0.54	-0.21	1.00						
CesA6	0.33	0.01	0.02	0.09	-0.15	1.00					
CesA7	0.70	0.21	-0.02	0.39	-0.13	0.29	1.00				
CesA8	0.63	0.34	-0.06	0.62	-0.38	0.22	0.60	1.00			
CesA10	0.30	0.03	-0.15	0.18	-0.25	0.13	0.41	0.24	1.00		
CesA11	0.32	0.09	-0.16	0.19	-0.27	0.16	0.45	0.31	0.93	1.00	
CesA12	0.33	0.02	-0.10	0.19	-0.23	0.19	0.51	0.27	0.89	0.85	1.00

All three of the secondary wall forming CesA proteins reported in Arabidopsis (IRX1, IRX3, and IRX5) have been reported to be involved in the formation of a functional cellulose synthase catalytic complex (Taylor e*t al., supra* (2003)). For the *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* genes, the correlation coefficients are around 0.9 among different pairs, indicating that these genes are mostly coexpressed.

A comparison between the expression levels of the *Zea mays CesA* genes in stem and leaf tissue was conducted using the MPSS expression data from Figure 5 and tabulated in Table 9. Suitable promoters for driving HCHL expression must show a significant tissue-specific expression pattern. Based on

the data provided in Table 9, it is clear that the promoters of the genes for *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* exhibit a suitable tissue-specific expression pattern. The respective promoter sequences were identified and are provided as SEQ ID NOs:81, 82, and 83.

Table 9 Comparison Between Expression Levels of Various *Zea mays CesA* Genes in Stem and Leaf Tissue Using MPSS

Gene Name	Leaf (ppm)	Stalk (ppm)	Stalk/leaf
CesA1	63	230	3.6
CesA2	0	0	0.0
CesA3	46	73	1.6
CesA4	8	30	3.7
CesA5	86	83	1.0
CesA6	262	179	0.7
CesA7	51	296	5.8
CesA8	63	284	4.5
CesA10	41	1033	25.0
CesA11	37	639	17.2
CesA12	16	370	22.8

EXAMPLE 6

Identification of *Oryza savita* Orthologs Using Maize Genes encoding the Cellulose Synthesis Catalytic Complex

The nucleic acid sequences for *ZmCesA10* (SEQ ID NO:31), *ZmCesA11* (SEQ ID NO:33), and *ZmCesA12* (SEQ ID NO:35) were used for a BLAST analysis against the rice BAC DNA (National Center for Biotechnology Information, Bethesda, MD) database. The results of the analysis, including the closest matching entry in the rice BAC database are listed in Table 10. Thus, the rice genome appears to contain three genes that are very closely related to *ZmCesA10*, *ZmCesA11*, and *ZmCesA12*, respectively. The nucleic acid sequences of the corresponding rice orthologs are set forth as SEQ ID NOs:37, 39, and 41, respectively. The corresponding deduced amino acid sequences are set for as SEQ ID NOs:38, 40, and 42, respectively.

Table 10

Sequence Analysis Results

Gene Name	Similarity Identified in Rice BAC Database (NCBI)	Identities[a]	Score (bits)	E-value[b]
ZmCesA10	gi|22711595|gb|AC022457.8	543/597 (90%)	381	0.0
ZmCesA11	gi|15146360|dbj|AP003237.2	487/524 (92%)	745	0.0
ZmCesA12	gi|21396530|dbj|AP005420.1	564/613 (92%)	827	0.0

[a] Identity is defined as percentage of nucleic acids that are identical between the two nucleic acid sequences.

[b] Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

EXAMPLE 7

Identification of Promoters from *Oryza savita* (japonica cultivar group) Genes Orthologous to *Zea mays ZmCesA10*, *ZmCesA11*, and *ZmCesA12* Genes

Based on sequence homology to the Arabidopsis genes *AtCesA8 (IRX1)*, *AtCesA7 (IRX3)*, and *AtCesA4 (IRX5)* and the tissue-specific expression pattern of the maize genes *ZmCesA10*, *ZmCesA11*, and *ZmCesA12*, it appears that these genes encode proteins involved in the formation of the cellulose synthesis catalytic complex catalytic responsible for cellulose deposition in the secondary cell walls of the vascular system of the corn stalk. In Example 6, the sequences of the maize genes were used to identify the sequences of the orthologous genes of rice. Also disclosed is the unexpected finding that gene function and the expression pattern of secondary cell wall-forming cellulose synthases is conserved between dicotyledonous plants (*Arabidopsis thaliana*) and monocotyledonous plants (*Zea mays*). This finding strongly suggests that the rice orthologues of *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* will have an expression pattern that is indistinguishable from those of their corn counterparts. Sequences set forth as SEQ ID NOs:43, 44, and 45 represent 2500 bp of rice genomic DNA sequence found immediately upstream (5') of the inferred start codon of the three genes (SEQ ID NOs:37, 39, and 41, respectively) that are orthologs of the *ZmCesA10*, *ZmCesA11*, and *ZmCesA12* genes, respectively. The sequences include putative regulatory elements such as cis-acting elements, transcription start sites and 5' UTRs of the rice genes. These sequences or part of these sequences can be used as promoters to target

expression of *HCHL* genes to the plant stalk as outlined in Examples 3 and 8. These promoters will be of particular use to target expression of *HCHL* genes in transgenic monocotyledonous plants such as sugar cane

EXAMPLE 8

Expression of HCHL in Plants Using Tissue-Specific Promoters

The isolation of the gene encoding the *Pseudomonas putida* DSM 12585 HCHL enzyme is described in Example 1. The methods for constructing plasmids for tissue-specific expression are described in Examples 2 and 3. Briefly, primer pairs can be chosen to amplify the suitable promoters from *Arabidopsis thaliana* and *Oryza savita* (japonica cultivar group), respectively. Genomic DNA from each respective source organism can be isolated using methods known in the art (Maniatis, *supra*). Primer pairs are chosen to amplify the respective genes from the genomic DNA (Table 11). The second member of the primer pair is designed to introduce a *Nco*I site (CCATGG) at the initiator methionine codon of the respective gene and facilitates generation of translational fusions of genes that contain *Pag*I (TCATGA) or *Nco*I sites at the start codon (Table 11). A pSKII+ plasmid containing a PCR-generated variant of the *HCHL* gene containing a *Pag*I site at the start codon is partially digested with *Pag*I and a *Pag*I/*Sst*I fragment is released from the vector by complete digestion with *Sst*I. In this example, the variant is created using the *Pseudomonas putida* DSM 12585 *HCHL* coding sequence (SEQ ID NO:5). However, methods to PCR-generate variants of genes so that a *Pag*I site is introduced at the initiator methionine codon for translational fusions is known in the art (Maniatis, *supra*). The respective promoter is released from the pCR2.1 vector by digestion with *Xba*I/*Pag*I. The respective promoter and the *HCHL* gene are assembled in a suitable plant transformation vector that has been digested with suitable restriction enzymes such as *Xba*I and *Sst*I in a three-way ligation. Plasmid DNA is used for agrobacterium-mediated transformation of Arabidopsis plants as previously described.

Table 11 Examples of Primer Pairs Suitable to Create Various Chimeric *HCHL* genes (based on *Pseudomonas putida* DSM 12585 HCHL)

Genomic DNA Source	Promoter	Primer Pair Member #1	Primer Pair Member #2
A. thaliana	*AtCesA4* (*IRX5*) (SEQ ID NO:46)	Primer 19 (SEQ ID NO:47)	Primer 20 (SEQ ID NO:48)
A. thaliana	*AtCesA8* (*IRX1*) (SEQ ID NO:49)	Primer 21 (SEQ ID NO:50)	Primer 22 (SEQ ID NO:51)
O. savita (japonica cultivar)	Ortholog of *Z. mays ZmCesA10* (SEQ ID NO:43)	Primer 23 (SEQ ID NO:52)	Primer 24 (SEQ ID NO:53)
O. savita (japonica cultivar)	Ortholog of *Z. mays ZmCesA11* (SEQ ID NO:44)	Primer 25 (SEQ ID NO:54)	Primer 26 (SEQ ID NO:55)
O. savita (japonica cultivar)	Ortholog of *Z. mays ZmCesA12* (SEQ ID NO:45)	Primer 27 (SEQ ID NO:56)	Primer 28 (SEQ ID NO:57)

Analysis of chimeric gene expression and kinetic analysis can be accomplished as described in Example 4.

EXAMPLE 9

Evaluation of Alternative HCHL Enzymes

Producing pHBA by HCHL in stalk tissue is limited by enzyme activity even if stalk-specific promoters are employed. Thus, further pHBA productivity improvements require the application of HCHL enzymes with better catalytic efficiency or the co-expression of several divergent HCHL enzymes that can be co-expressed without triggering transcriptional or posttranscriptional gene silencing. A BLAST search of the public domain databases for putative HCHL enzymes was conducted. Figure 6 shows a phylogenetic tree of a CLUSTAL W alignment of putative and *bona fide* HCHL enzymes in public databases. With the exception of a putative HCHL enzyme of *Caulobacter crescentus*, the name of the other potential HCHL enzymes are not provided since their catalytic activities have not been investigated. Figure 6 illustrates that a large source of divergent "HCHL-like" enzymes that could be exploited for further improvements of pHBA accumulation in plants. The putative HCHL enzyme of *Caulobacter crescentus* shares only 54 % amino acid identity to the HCHL enzymes from

Pseudomonas putida and *Pseudomonas fluorescens* AN103 based on BLAST analysis.

Expression Cloning of *HCHL* Gene of *Caulobacter Crescentus*

Genomic DNA of the *Caulobacter crescentus* strain used for the genome sequencing project (Nierman *et al.*, *PNAS*, 98(7):4136-4141 (2001)) was obtained from ATCC and used for PCR amplification of the *HCHL* ORF using the Primer 29 : CCAAGGACCGCATATGACAGACGCCAACGAC (SEQ ID NO:68) and Primer 30: CCTCCCCCTCGCAAGCTTTCAGCTCTGCTTGG (SEQ ID NO:69). The primers introduce *Nde*I and *Hind*III sites flanking the ORF. The PCR product was digested with *Hind* III and *Nde*I and ligated to the pET29a vector DNA that had been cut with the same restriction enzymes. Recombinant plasmid DNA was sequenced and introduced into BL21DE3 cells.

Expression Cloning of *HCHL* Gene of *Pseudomonas fluorescens* (AN103)

To evaluate the utility of the Caulobacter HCHL enzyme for producing pHBA in plants, it was important to compare its kinetic properties to those of the two Pseudomonas enzymes that have been previously utilized to produce pHBA in plants. The Applicants cloned, expressed, and purified the HCHL enzyme of *Pseudomonas fluorescens* (AN103). Plasmid DNA of pSP72 (Promega) containing the *Pseudomonas fluorescens* (AN103) *HCHL* ORF is described in Mayer *et al.* (*supra*). It was used for PCR amplification of the *HCHL* ORF using the Primer 31: GAGAGCATCCATATGAGCACATACGAAGGTCGC (SEQ ID NO:70) and Primer 32: CGCAGCGTCAAGCTTCAGCGTTTATACGCTTGC (SEQ ID NO:71). The primers introduce *Nde*I and *Hind*III sites flanking the ORF. The PCR product was digested with *Hind* III and *Nde*I and ligated to the pET29a vector DNA that had been cut with the same restriction enzymes. Recombinant plasmid DNA was sequenced and introduced into BL21DE3 cells.

Recombinant Production, Purification, and Analysis of Kinetic Properties.

HCHL enzymes were purified from cell-free extracts of BL21DE3 cells expressing the pET29a expression constructs by chromatography on Q-sepharose and hydroxyapatide as described in Example 1 for the native HCHL enzyme from *Pseudomonas putida* (DSM 12585). The following calculated properties of the HCHL proteins were used to determine kinetic properties of the HCHL enzymes.

HCHL *Caulobacter crescentus*: Molecular weight: 31104.09, Molar extinction coefficient: 59690.

HCHL *Pseudomonas fluorescens* AN103: Molecular weight: 31007.39, Molar extinction coefficient: 50190.

The enzyme preparations used to determine the kinetic properties were analyzed by Coomassie staining of PAGE gels, indicating that the both enzymes were at least 90% pure.

Figure 7 and Table 12 summarize kinetic properties of the HCHL enzymes with the pHCACoA substrate. They were determined in standard HCHL enzyme reactions by using 1.4, 2.6, and 0.8 ng of purified HCHL enzymes of *P. putida*, *P fluorescens*, and *C. crescentus*, respectively. pHCACoA concentrations were varied from 0.9 to 440 μM. The high turnover number of the HCHL enzyme of *Pseudomonas fluorescens* (AN103) was more than four times higher than the kcat reported by Mitra *et al.* for the same enzyme (*Arch. Biochem. Biophys.*, 365(1):6-10 (1999)). The HCHL enzyme of *C. crescentus* unexpectedly showed a 50 % improvement of catalytic efficiency (Kcat/Km) when compared to the *Pseudomonas fluorescens* AN103 enzyme. Thus, the Caulobacter HCHL protein provides an ideal candidate for a catalyst to achieve further improvements of pHBA productivity in plants.

Table 12. Kinetic activity comparison between various HCHL enzymes

HCHL Enzyme Source	Km (μM)	Vmax (nkat mg^{-1})	Kcat (s^{-1})	Kcat/Km
P. putida	2.4	43	3.4	1.41
P. fluorescens	3.8	157	9.7	2.55
C. crescentus	4	240	15.2	3.8

Constitutive Expression of *Pseudomonas fluorescens* AN103 and *Caulobacter crescentus HCHL* Genes in Plants

Transgenic lines can be generated that express HCHL enzyme from *Pseudomonas fluorescens* and *Caulobacter crescentus* under the control of constitutive promoters. This should be considered as a first step to investigate whether improved kinetic properties of the HCHL enzymes of Caulobacter result in higher levels of accumulated pHBA in stalk tissue when compared to Pseudomonas HCHL enzymes.

Construction of a Vector for Expression of HCHL *Caulobacter crescentus* in Transgenic Plants

To generate a construct for constitutive expression of the *Caulobacter crescentus* HCHL enzyme in transgenic plants, a 0.9 kb *Xba*I/*Hind*III DNA fragment (containing the full-length *HCHL* Caulobacter ORF and 42 bp of 5' untranslated DNA (derived from the pET29A vector) immediately upstream of

the initiation codon) was excised from the pET29a construct used for recombinant enzyme production and cloned into the pGEM3zf+ vector (Promega). This cloning step introduces a *Bam*HI site upstream of the Caulobacter *HCHL* start codon. Recombinant pGEM3zf+ DNA containing the *HCHL* gene was linearized by digestion with *Hind*III. Linearized plasmid DNA was purified and overhanging DNA ends were filled-in with T4 DNA polymerase (New England Biolabs, MA, USA) according to manufacturer instructions. The *HCHL* gene was released from the plasmid by digestion with *Bam*HI. The restriction fragment was ligated to *Bam*HI and *Hpa*I digested pBE856 DNA. This resulted in replacement of the *FlpM* recombinase ORF in pBE856 with the *HCHL* gene of Caulobacter, situated between the constitutive *SCP1* promoter and 3' untranslated region of the potato proteinase inhibitor II (*PIN II*) gene. The resulting binary vector, HCHL Caulobacter expression construct was used for plant transformation as described in General Methods. Plasmid pBE856 (*SCP-FlpM*) was previously constructed by cloning a 2172 bp *Xba*I - *Eco*RI fragment containing a chimeric *SCP1:FlpM:3' Pin* gene into the multiple cloning site of the binary vector pBE673 (described below), after cleavage of the latter with *Xba*I and *Eco*RI.

The *SCP1:FlpM:Pin* gene is comprised of a synthetic *35S* promoter (*SCP1*) (Bowen *et al.*, U.S. (2000), 31 pp., Cont.-in-part of U.S. Ser. No. 661,601, abandoned. CODEN: USXXAM US 6072050 A 20000606), which is fused at its 3' end to the ORF of the *FlpM* recombinase, which is fused at its 3' end to the 3' *PIN* region derived from the *Solanum tuberosum* proteinase inhibitor II gene (GenBank® Accession No. L37519). Plasmid pBE673 was derived from pBin 19 (GenBank® Accession No. U09365) by replacing an 1836 bp *Bsu*36a-*Cla* I fragment of pBin19 (which contains the 3' end of the nopaline synthase (*nos*) promoter, the *npt II* (kanamycin resistance) ORF, and the 3' *nos* region) with a 949 bp *Bsu*36I-*Cla* I fragment (which contains (5' to 3'): a 106 bp fragment comprising the 3' end of *nos* promoter (nucleotides 468-574 described in GenBank® Accession Nos. V00087 and J01541; see also Bevan *et al.*, *Nucleic Acids Res.*, 11 (2), 369-385 (1983)), a 5 bp GATCC sequence, a 551 bp fragment corresponding to the *Streptomyces hygroscopicus* phosphothricin acetyl transferase (basta resistance) ORF (GenBank® Accession No. X17220) except that the termination codon was changed from TGA to TAG, an 8 bp TCCGTACC sequence, and a 279 bp 3' *nos* region (nucleotides 1824-2102 of GenBank® Accession Nos. V00087 and J01541 described above)).

Vector for Expression of HCHL *Pseudomonas fluorescens* (AN103) in Transgenic Plants

The binary vector plasmid for expressing the *HCHL* gene of *Pseudomonas fluorescens* (AN103) in transgenic plants is described in detail by Mayer *et al.* (*Plant Cell*, 13:1669-1682 (2001)). Both binary vectors were introduced into Arabidopsis plants by agrobacterium-mediated transformation. Transgenic lines carrying the *HCHL* gene of *P. fluorescens* and *C. crescentus* were selected on kanamycin and phosphinotrine, respectively, and grown in soil for eight weeks. pHBA concentration was determined in basal stem segments. Table 13 shows that pHBA levels are significantly higher in the Caulobacter HCHL transgenics in comparison to the Pseudomonas HCHL transgenics. In the best Caulobacter HCHL transgenics, pHBA levels in the basal stem segments are nearly doubled.

Whole stalk material was harvested after ten weeks and subjected to pHBA analysis. This analysis confirmed our previous observation indicating that a new high threshold of pHBA accumulation in whole stalk tissue of nearly 50 mg/g DW (dry weight) could be established by expression of the Caulobacter *HCHL* gene under control of the constitutive *SCP1* promoter.

T2 plants of the Caulobacter and *Pseudomonas fluorescens* HCHL transgenics were germinated on selective media and grown in soil for 6 weeks to obtain sufficient stalk tissue for analysis of HCHL enzyme activity. Table 14 shows that expression of the Caulobacter *HCHL* gene led to an increase of specific HCHL activity in stalk tissue when compared to the HCHL Pseudomonas transgenics that reflects the differences in kinetic properties between the two enzymes that were detected *in vitro.*

Table 13. pHBA levels measured in several HCHL transgenics

Construct	n	Basal stalk pHBA average (mg/g FW)	Basal stalk pHBA highest (mg/g FW)
35S HCHL P. fluorescens	42	1.72	6.0
SCP1 HCHL C. crescentus	72	2.4	11.8

Table 14. HCHL specific activity measured between various constructs

Construct	Line	Rate (pkat/mg protein)	pHBA (mg/g DW)
35S HCHL P. putida	276	160	19.8
35S HCHL P. fluorescens	374	480	30.0
SCP1 HCHL C. crescentus	10	614	49.2
SCP1 HCHL C. crescentus	24	610	47.4
SCP1 HCHL C. crescentus	29	653	46.0

Data in Tables 13 and 14 show that the higher catalytic efficiency of the HCHL enzyme of *Caulobacter crescentus* compared to HCHL enzymes of Pseudomonas is responsible higher specific HCHL activity and higher levels of pHBA accumulation in transgenic plants. An alternative explanation for this observation, however, may lie in the nature of the constitutive promoters that are expressing the respective *HCHL* genes. The Pseudomonas genes are expressed under the control of the double enhanced *35S* promoter. The *HCHL* gene of Caulobacter, on the other hand, is expressed under the control of the *SCP1* promoter. Although both promoters are ultimately derived from the *35S* promoter, the promoters may differ in the level of gene expression that they can confer. Thus, the higher levels of HCHL activity and pHBA accumulation of the Caulobacter HCHL transgenics may merely reflect higher transcript levels that are achieved with the *SCP1* promoter. In order to investigate this further, seed material of lines 374 and 29 were germinated on MS media containing glufosinate. Herbicide-resistant plants were transferred to soil and grown for 8 weeks. Stalk tissue was harvested and subjected to RNA isolation using standard procedures (Maniatis, *supra*) and HCHL enzyme activity was measured. HCHL transcript levels in line 374 and 29 were detected by real time PCR as follows:

Real time RT-PCR data was generated on an ABI 7900 SDS instrument (Applied Biosystems, CA, USA). Dual labeled Taqman probes and RT-PCR primers were designed for all mRNA targets using ABI Primer Express v 2.0 software package (Applied Biosystems, CA, USA) using default settings. The

probes were labeled at the 5' end with the reporter fluorochrome 6-carboxyfluorescein (6-FAM) and the quencher fluorochrome 6-carboxy-tetramethyl-rhodamine (TAMRA) at the 3' end. Real Time one step RT-PCR reactions were set up using 1 μM final concentration of both the forward and reverse RT-PCR primers, 250 nm final concentration of the Taqman probe, 5U ABI Multiscribe Reverse transcriptase, 8 U ABI RNAse Inhibitor, and 10 μL ABI Taqman Universal PCR Master Mix. The reaction volume was adjusted to 19 μL with RNase free water and 1 μL RNA was added at concentrations of 50 to 0.78 ng/μL. Reverse transcription was carried out for 30 min at 48 °C followed by 10 min at 95 °C for AmpliTaq Gold activation. Real time data (Cycle threshold or "Ct's") was collected during 40 cycles of PCR; 95 °C, 5 sec, 60 °C, 1 min.

Actin Real Time Data

Real time RT-PCR data were generated using a set of primers and probes targeting the *ACTIN2* gene of Arabidopsis (GenBank® Accession No. U41998) which has been shown to be constitutively expressed (An *et al., Plant J.*, 10 (1):107-121 (1996)).

The following primers were used

Primer 33 (Actin2RT-FWD) (SEQ ID NO:72): TGA GAG ATT CAG ATG CCC AGA A

Primer 34 (Actin2RT-REV) (SEQ ID NO:73): TGG ATT CCA GCA GCT TCC AT

Primer 35 (Actin2Probe) (SEQ ID NO:74): TCT TGT TCCA GCC CTC GTT TGT GGG

The objective was to identify and normalize RNA concentration differences between the samples isolated from the Caulobacter HCHL transgenic (29) and the Pseudomonas HCHL transgenic (374). The real time data for 25 ng, 12.5 ng, and 6.25 ng total RNA is shown in Table 15. It lists threshold cycle (Ct) determined for both RNA samples. The Ct value identifies the PCR cycle number at which the reporter dye emission intensity rises above background noise. The Ct value is determined at the most exponential phase of the PCR reaction and is therefore a more reliable measure of PCR target concentration than end-point measurements of accumulated PCR products in conventional reverse transcriptase-PCR experiments. The Ct value is inversely proportional to the copy number of the target template. The mean Ct values of three independent analyses are shown; corresponding SD values are also indicated. Both RNA preparations show very similar Ct values for each of the three concentrations. The % difference between the two vary from 0.0 % to 0.3 %. Since the *ACTIN2* gene is constitutively expressed this data indicates that the RNA samples are of very similar concentration. The actin real time PCR

data was used to normalize the real time expression data for the *HCHL* genes shown below.

Table 15. Real time PCR analysis comparing the threshold cycles of the *ACTIN2* control used for normalization.

	Ct Values	Ct Values	
ng RNA	374 Pseudomonas Actin	29 Caulobacter Actin	% Difference
25	19.38 ± 0.05	19.38 ± 0.12	0.0%
12.5	20.34 ± 0.10	20.37 ± 0.12	0.1%
6.25	21.97 ± 0.31	22.03 ± 0.20	0.3%

HCHL Real Time Data

Real time RT-PCR data was generated using primers and probes designed specifically for the Pseudomonas or the Caulobacter *HCHL* gene. The following primers were used:

Primer 36 (HCHL CAUL RT-FWD) (SEQ ID NO:75): GCC TGG GTG AAG TTC AAT CG

Primer 37 (HCHL CAUL RT-REV) (SEQ ID NO:76): CCA TCA TGC GAC GGT TCA G

Primer 38 (HCHL CAUL Probe) (SEQ ID NO:77): CCC GAT AAG CGC AAC TGC ATG AG

Primer 39 (HCHL PFL RT-FWD) (SEQ ID NO:78): TGC GCC GAC GAA GCA

Primer 40 (HCHL PFL RT-REV) (SEQ ID NO:79): GTT GCC CGG CGG GAT A

Primer 41 (HCHL PFL Probe) (SEQ ID NO:80): TTC GGT CTC TCG GAA ATC AACTG

The PCR efficiency of these two different RNA-primer sets was compared based on how the Ct values changed across the entire range of Arabidopsis RNA dilutions from 50 to 0.78 ng/reaction (rxn). Linear regression analysis of the obtained Ct values versus the log of the RNA concentration was performed. The slopes of the two sets of data were used to calculate the RT-PCR efficiency for both sets of RT-PCR primers and probes. The calculation was performed as described (Pfaffl, M.W., *Nucleic Acids Res.*, 29(9):e45 (2001)). The data is shown in Table 16. RT-PCR efficiency for the Caulobacter and Pseudomonas HCHL primers and probe is 1.96 and 1.94, respectively; 2.0 is the theoretical maximum efficiency for exponential amplification in a PCR reaction. The efficiencies are very similar. Therefore, the real time data acquired with the

HCHL specific primers and probes can be directly compared. The actin data (Table 15) were used to normalize for differences in the RNA concentration of both RNA samples.

Table 16. Comparison of Real Time RT-PCT Efficiency.

ng RNA	Log ng RNA	29 Caulobacter Ct Values[a]	374 Pseudomonas Ct Values[a]
50.00	1.70	17.14 ± 0.02	15.91 ± 0.02
25.00	1.40	18.27 ± 0.03	17.00 ± 0.05
12.50	1.10	19.33 ± 0.05	17.92 ± 0.06
6.25	0.80	20.27 ± 0.06	19.07 ± 0.06
3.13	0.49	21.35 ± 0.00	20.11 ± 0.04
1.57	0.19	22.35 ± 0.05	21.16 ± 0.04
0.78	-0.11	23.38 ± 0.05	22.16 ± 0.04
	Slope:	-3.43	-3.47
	Correlation Coefficient (R^2)	1.00	1.00
	RT-PCR Efficiency[b]:	1.96	1.94

[a] Values represent the mean of n=3 replicates, ± = SD
[b] Efficiency = ($10^{(-1/slope)}$); 2.0 is maximum value for exponential amplification

Relative Expression in Arabidopsis of the Pseudomonas and Caulobacter *HCHL* gene

The real time data from the tables above was used to calculate the expression of the Caulobacter *HCHL* gene relative to the Pseudomonas *HCHL* gene (Pfaffl, M. W., *supra*). The relative expression data is shown in Table 17 for three different dilutions of the Arabidopsis RNA preps. The data indicate that for every mRNA transcript of the Pseudomonas HCHL gene that is produced only 0.40 – 0.46 Caulobacter transcripts are produced in the equivalent amount of Arabidopsis tissue.

Table 17. Relative Expression of Arabidopsis RNA

ng RNA in RT-PCR	HCHL Efficiency	Actin Efficiency	CT HCHL* (Pseudo - Caulo)	CT Actin* (Pseudo - Caulo)	Relative Expression Caulobacter relative to Pseudomonas
25ng RNA	1.95	1.7	-1.27 ± 0.06	0 ± 0.13	0.43 ± 0.06
12.5ng RNA	1.95	1.7	-1.41 ± 0.07	-0.03 ± 0.15	0.40 ± 0.07
6.25ng RNA	1.95	1.7	-1.2 ± 0.09	-0.06 ± 0.37	0.46 ± 0.09

*Difference of means of n=3 replicates; ± = 1SD

HCHL enzyme activity

The tissue used for RT-PCR experiments was also subjected to assays of HCHL activity. Table 18 shows that specific HCHL enzyme activity in stem tissue line 29 is 26 % higher than in line 374. Real time PCR experiments revealed that HCHL transcript levels in lines 29 are lower than those detected in 374. Thus, strong evidence is provided for the conclusion that enhanced HCHL enzyme activity and pHBA accumulation observed in transgenic plants expressing the *HCHL* gene of *Caulobacter crescentus* is due to in improved kinetic properties of the HCHL enzyme.

Table 18. Comparison of HCHL enzyme activity in stem tissue of various constructs

Construct	Line	HCHL rate (pkat/mg protein)
35S HCHL P. fluorescens	374	254 +/- 9
SCP1 HCHL C. crescentus	29	320 +/-2

The *HCHL* gene from *Caulobacter crescentus* (with prior undisclosed activity) shows a 50 % improvement of catalytic efficiency (Kcat/Km) when compared *in vitro* to a Pseudomonas HCHL enzyme described in the literature. Expression of this *HCHL* gene in transgenic plants resulted in increased pHBA accumulation in stalk tissue from 3 % DW (observed with *HCHL* gene from Pseudomonas) to 4.9 % DW. Transgenic plants expressing the *HCHL* gene of

Caulobacter under control of constitutive promoters exhibited detrimental phenotypes similar to those observed when *HCHL* genes of Pseudomonas were expressed in transgenic plants. These phenotypes included delayed development, depletion of soluble phenylpropanoids (sinapoyl malate) in leaf tissue and early senescence in leaf tissue. However, as described in Example 3 of this application, these negative side effects can be avoided through expression of *HCHL* genes under the control of tissue-specific promoters; specifically promoters of cellulose synthase genes that represent *AtCesA8 (IRX1)*, *AtCesA7 (IRX3)*, and *AtCesA4 (IRX5)* or promoters of orthologous genes present in other plant species.

The low level (<57%) of sequence identity of *HCHL* genes of *Pseudomonas putida* (DSM12585) and *Pseudomonas fluorescens* AN103 relative to the *HCHL* gene of *Caulobacter crescentus* enables co-expression of both *HCHL* genes in a single plant cell. This elegant route to even higher levels of *HCHL* gene expression in plant cells avoids co-suppression problems that would arise from co-expression of closely-related *HCHL* genes in plants.